格 致 方 法 | 社会科学研究方法系列

整体网分析

UCINET软件实用指南（第二版）

Lectures on Whole Network Approach

A Practical Guide to UCINET (2nd)

刘军 编著

格致出版社 上海人民出版社

格 致 方 法 | 社会科学研究方法系列

整体网分析

UCINET软件实用指南（第二版）

Lectures on Whole Network Approach

A Practical Guide to UCINET (2nd)

刘军 编著

格 致 出 版 社 上海人民出版社

图书在版编目(CIP)数据

整体网分析:UCINET软件实用指南/刘军编著.—2版.—上海:格致出版社:上海人民出版社,2014
(格致方法.社会科学研究方法系列)
ISBN 978-7-5432-2398-1

Ⅰ.①整… Ⅱ.①刘… Ⅲ.①社会关系-研究-应用软件-指南 Ⅳ.①C912.3-39

中国版本图书馆CIP数据核字(2014)第120167号

责任编辑 王亚丽
装帧设计 人马艺术设计·储平

整体网分析(第二版)——UCINET软件实用指南
刘 军 编著

出 版	世纪出版股份有限公司 格致出版社 世纪出版集团 上海人民出版社 (200001 上海福建中路193号 www.ewen.cc) 格致出版 编辑部热线 021-63914988 市场部热线 021-63914081 www.hibooks.cn	印 刷	上海商务联西印刷有限公司
		开 本	787×1092 1/16
		印 张	23
		插 页	2
		字 数	378,000
		版 次	2014年8月第1版
发 行	上海世纪出版股份有限公司发行中心	印 次	2014年8月第1次印刷

ISBN 978-7-5432-2398-1/C·104 定价:48.00元

国家社会科学基金项目（批准号：13BSH054）
哈尔滨工程大学中央高校基本科研基金项目（编号：HEUCF20141302）
国家社会科学基金重大招标项目（批准号：13&ZD177）

刘 军

男，1970年生于黑龙江省桦南县，2003年毕业于北京大学社会学系，获博士学位，曾于2002年赴英国Lancaster大学和Essex大学社会学系做访问学者，于2010—2011年赴美国做访问学者。现任哈尔滨工程大学人文社会科学学院社会学系主任、教授、博士生导师，中国社会学会常务理事，中国社会学会社会网专业委员会副会长。研究专长为社会网络分析、社会学方法论。

前　言

社会网研究在上世纪 90 年代之后逐渐崛起，成为一个新的社会学研究领域，并吸引了大量的学者加入此研究队伍。在美国社会学界，社会网研究被认为与理性选择学派、新制度论学派并列，它不但是未来最具影响力的三大学派之一，而且其影响力还扩散到其他领域。在管理学界，它成为组织行为研究、战略研究、知识传播与创新研究以及消费者行为研究的新研究范式。过去几年，每年在管理学最好的期刊 ASQ、AMJ 以及 AMR 上，平均会刊登 20 篇以上的相关论文，并且在 2002 年时，社会网及社会关系还成为美国管理学年会的主题。在地理学研究领域，它被用来研究产业集群及产业区位问题，在政治学研究领域，它被用来研究投票行为及政策形成过程，在经济学研究领域，则被用来分析组织治理及交易治理问题。社会网研究无疑已成为跨学科的一个重要研究领域。

另一个重大的发展则是在世纪之交，小世界网络(small-world network)的研究在物理学界及社会学界中被同时提出，之后便迅速在各个领域蔓延，不但在社会科学及管理学中有了丰硕的研究成果，而且在万维网、电话网、电力网、脑神经网以及生态环境上也都有着突飞猛进的进步。随之兴起的是复杂性(complexity)研究，与社会网相结合，在收入分配、产业集群、供应网络、知识传播网络以及战略联盟网络等议题上都建立了新的研究范式，这个研究范式为网络的演化问题开启了一扇窗，可以研究人类自我组织(self-organization)的过程，从而奠定了网络动态学(network dynamics)的研究基础。

自我组织、网络结构自我演化、自我寻找社会秩序，正好标志着中国社会的特质以及中国人无为而治的管理哲学，所以，这一套研究范式正是用来研究本土社会学及本土管理学的最佳范式。目前，国内研究者日益增多，其最主要的相关学会——中国社会学会社会网暨社会资本研究专业委员会——每年在会议中可以收到百篇以上的论文，在国内知名期刊上出版数个专辑，而且成立短短两年半以来，已经吸引了 400 个研究同好的加入，在国内社会学、管理学界成为一股新的研究力量。

然而，想进入社会网研究的学者及学生面对的却是一道技术上的难关，社会网的分析技术经过 30 多年的发展已渐臻完备，往往需要有不错的数学功底才能

完全掌握,尤其要有图论及代数的基础方能胜任分析之责。沃瑟曼和福斯特(Wasserman & Faust, 1994)总结了这些社会网分析的数学工具,写下了洋洋洒洒一本厚书,堪称社会网分析技术的百科全书。这本书在数学推导上虽然严谨,却不易读懂,一般对数学不算太好的文科及管理方面的研究者而言,略显艰难。这些分析技巧其实已经有了容易上手,弹指之间即可算出答案的软件,如UCINET,这本书没有介绍,更缺乏社会网理论的引介,所以对多数读者而言,它是一本数学工具书,而不是一本入门的书。

约翰·斯科特(John Scott, 2001)写了一本社会网研究的入门书可以补其不足,其对社会网理论及研究的介绍堪称全面而易懂,然而它却不曾包括任何社会网分析技术的细节,所以读者看懂了社会网研究路数后,还是无法着手作一些研究。

鉴于这些缺点,我曾经写了一本《社会网分析讲义》,将社会网理论、研究方法以及社会网分析的技巧一起进行了介绍,而且鉴于社会网分析数学的艰涩难懂,也同时介绍了UCINET的使用方法。然而这本书旨在引人入门,重点在于理论与方法的结合,并以研究范例进行说明,对社会网分析的介绍十分少,也不全面。书成之日,就觉得意犹未尽,总想写一本十分全面的书,再加上UCINET的使用方法,这样,既可以让读者一窥社会网分析技术之全豹,又能轻松上手。

去年初得到机会看到刘军教授这本书的初稿,方知此心愿可以罢唱,因为这本书既十分全面地介绍了社会网分析,又添加上了一些复杂网络的分析,如小世界的测量,更巨细靡遗地将UCINET的使用步骤一步一步地教会读者,可以说本书是国内目前学习整体网研究最全面又最容易上手的工具书。

刘军教授在国内推广社会网研究上一直不遗余力,不但常常开课教授社会网分析的课程,之前也曾经写过其他的工具书,还翻译了约翰·斯科特的书。这些社会网研究的入门介绍,再加上这本全面又易懂的分析技术工具书,初学的读者们一定可以很轻松地跨入社会网研究的门槛。

罗家德

笔于清华园

罗家德,2005,《社会网分析讲义》,北京:社会科学文献出版社。

Scott, John. *2000*. *Social Network Analysis: A Handbook*. London: Sage Publications Ltd.

Wasserman, Stanley and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

目　录

第一章　社会网络分析简介 …… 001

第一节　社会网络分析的含义 …… 001

一、社会网络的含义 …… 001

二、1-模网和 2-模网 …… 005

第二节　用图和矩阵表示社会网络 …… 006

一、社会网络的图形表达 …… 007

二、社会网络的矩阵表达 …… 009

第三节　社会网络研究的范围和理论贡献 …… 011

一、研究范围 …… 011

二、理论贡献 …… 014

第二章　整体网研究概要 …… 016

第一节　整体网研究的内容 …… 016

一、整体网的分类 …… 017

二、整体网的构成和规模 …… 017

三、整体网的密度 …… 019

四、整体网成员之间的距离 …… 020

五、整体网的结构研究 …… 022

第二节　整体网资料的收集 …… 028

一、整体网资料的收集方法 …… 028

二、问卷设计应注意的问题 …… 036

三、整体网资料的整理和分析 …… 037

四、整体网资料的相关问题 …… 038

第三节　整体网研究的优势及局限 …… 038

一、整体网研究的优势 …… 039

二、整体网研究的局限 …… 039

第三章　整体网分析工具——UCINET 简介 …… 042
第一节　UCINET 的运行环境 …… 042
一、UCINET 的数据分析格式 …… 043
二、UCINET 的主菜单功能简介 …… 051
三、UCINET 中矩阵运算的句法 …… 070
第二节　UCINET 数据的输入 …… 075
一、初始数据文件 …… 075
二、Excel 文件数据 …… 076
三、数据语言类型的文件 …… 076
四、UCINET 数据表编辑器 …… 094
第三节　UCINET 数据的预处理 …… 095
一、抽取子图和子矩阵 …… 095
二、数据的合并 …… 096
三、置换与排序 …… 097
四、转置和改型 …… 098
五、重新编码 …… 099
六、线性转换 …… 100
七、对称化处理 …… 100
八、捷径距离和可达性 …… 100
九、矩阵的组合 …… 101
十、标准化处理 …… 102
十一、改变模态 …… 103
第四章　均衡性分析 …… 106
第一节　符号图中的均衡研究 …… 107
一、社会心理学中的均衡研究 …… 107
二、有关均衡的模拟研究 …… 109
三、整体均衡研究 …… 111
第二节　有向关系中的均衡 …… 116
一、有向 P-O-X 三角关系 …… 116
二、正—负关系的均衡性分析 …… 117
三、有向三方组的谱系 …… 121

第五章　中心性和影响力…… 126
第一节　度数中心性…… 127
一、点的度数中心度…… 127
二、图的度数中心势…… 128
第二节　中间中心性…… 129
一、点的中间中心度…… 129
二、图的中间中心势…… 130
三、线的中间中心度…… 131
四、针对中间中心度的等级嵌套分析…… 133
第三节　接近中心性…… 134
一、点的接近中心度…… 134
二、图的接近中心势…… 135
三、多种中心度指数的同时计算…… 136
第四节　影响力分析…… 136
一、影响力的含义及测量…… 137
二、影响力指数计算示例…… 142
第五节　群体中心度…… 144
一、三类群体中心度指数的含义…… 144
二、群体中心度指数的计算举例…… 145
第六章　凝聚子群分析…… 153
第一节　凝聚子群概说…… 154
一、社会结构的质性分析和量化研究…… 154
二、凝聚子群的类型…… 155
第二节　基于互惠性的凝聚子群…… 157
一、派系的定义…… 157
二、对派系概念的评价…… 161
第三节　基于可达性的凝聚子群…… 162
一、n-派系…… 163
二、n-宗派…… 165
第四节　基于度数的凝聚子群…… 166
一、k-丛…… 166

二、k-核 …… 168
第五节 基于"子群内外关系"的凝聚子群 …… 170
一、成分 …… 171
二、块 …… 171
三、LS 集合 …… 172
四、Lambda 集合 …… 176
五、社会圈 …… 178
第六节 凝聚子群的分派指数 …… 180
一、单类网络中的分派指数 …… 180
二、多类网络中的分派指数 …… 183
第七节 凝聚子群分析示例 …… 189
一、进行凝聚子群分析的几个指导性步骤 …… 190
二、凝聚子群分析示例 …… 195
第七章 社会网络的关联性与凝聚度 …… 202
第一节 关联度和凝聚力 …… 202
一、根据社会关系界定的凝聚力 …… 203
二、根据独立途径数界定的凝聚力 …… 204
第二节 关联度的测量 …… 206
一、关联度 …… 207
二、图的等级度 …… 209
三、图的效率 …… 210
四、最近上限 …… 211
第三节 关联性与小世界 …… 214
一、小世界的含义 …… 215
二、小世界的测量指标 …… 216
三、小世界的测量示例 …… 218
第八章 块模型分析 …… 221
第一节 块模型的含义 …… 221
一、块模型 …… 222
二、CONCOR 方法 …… 230
第二节 块模型的应用举例 …… 234

第九章　个体网研究:结构洞与中间人 …… 237
第一节　局部桥和结构洞 …… 237
一、局部桥的含义及测量 …… 237
二、结构洞 …… 240
第二节　中间人分析 …… 254
一、中间人的分类 …… 255
二、中间人次数的计算 …… 256
第三节　其他个体网指标的计算 …… 260
一、个体网的指标 …… 260
二、个体网的同类性 …… 262
三、个体网的构成 …… 262
四、诚实中间人指数 …… 262
第十章　一致性分析 …… 264
第一节　一致性分析的含义 …… 264
第二节　一致性分析举例 …… 268
第十一章　2-模网研究 …… 273
第一节　2-模数据的矩阵分析和图形分析 …… 273
一、2-模网络向 1-模矩阵的转化 …… 274
二、二部“2-模图”分析 …… 281
第二节　2-模数据的定量分析 …… 284
一、2-模数据的中心性分析 …… 284
二、2-模数据的奇异值分解 …… 290
三、2-模数据的因子分析 …… 295
四、2-模网络的对应分析 …… 301
第三节　2-模数据的核心—边缘分析和分派分析 …… 304
一、2-模数据的核心—边缘分析 …… 305
二、2-模数据的分派分析 …… 306
第十二章　关系数据的假设检验 …… 309
第一节　点的属性层次的假设检验 …… 310
一、回归分析 …… 310
二、t 检验 …… 314

三、方差分析 …… 316
第二节 点—关系混合层次的假设检验 …… 318
一、类别变量与关系变量之间关系的检验 …… 318
二、连续变量与关系变量之间关系的检验 …… 326
第三节 关系—关系层次的假设检验 …… 330
一、QAP 的含义 …… 331
二、QAP 相关分析 …… 335
三、QAP 关系列联表分析 …… 336
四、QAP 回归分析 …… 337
第四节 密度检验 …… 340
一、一个网络的密度是否等于某个值 …… 340
二、两个网络的密度是否相等 …… 342
参考文献 …… 344
后记 …… 357

第一章

社会网络分析简介

中国社会是关系社会，社会学就是一门研究社会关系的学科。在当代中国，社会关系紧张，社会矛盾、社会冲突频发。在社会学恢复重建30多年的今天，研究社会关系的属性、结构和演变对于描述和解释中国社会的发展、变迁和创新社会治理模式具有重要意义。不过，国内学者对社会关系的研究（无论是量化研究还是质性研究）依然薄弱。国外有关社会关系的定量研究（社会网络研究）已经有近80年时间（刘军，2004：第二章）。尽管如此，“在过去的30多年中，经验社会研究主要由抽样调查（sample survey）控制着。但是，如人们常常指出的那样，调查是一种社会学的绞肉机（meat grinder），它把个体从其所在的社会情境（social context）中抽离出来，并确保研究对象之间不存在联系”（转引自Freeman，2004：1）。这是40多年前（即1968年）一位学者的看法，当时是正确的，“并且至今仍然是正确的”（Freeman，2004：2）。抽离个体所在的社会情境，不关注社会行动者之间的联系，这样的研究不能揭示社会现象的关系本质。

社会关系是复杂的，我们对各类社会关系既可以进行质性研究，又可以进行量化探讨。就后者来说，社会网络分析（social network analysis，SNA）就派上了用场，它拥有独特的本体论、认识论和方法论意涵，有力地扩展了社会结构研究。

第一节 社会网络分析的含义

本节简要介绍社会网络分析的基本概念。至于更丰富的内容请读者参见其他相关资料（Wasserman & Faust，1994；刘军，2004；罗家德，2010；沃瑟曼、福斯特，2012）。

一、社会网络的含义

一般认为，“网络”是由节点及节点之间的某种关系构成的集合，“社会网络”是由作为节点的社会行动者（social actor）及其间的关系构成的集合。不过在我

们看来,如果将上述"社会网络"界定中的"行动者"改为能动者(agent),会更恰当地表达社会网络的含义。也就是说,我们认为"社会网络"是作为节点的社会能动者(social agent)及其间的关系的集合。国内外的社会网络研究关注的多数是"行动者"(actor)而非"能动者",二者是不同的。前者指的是一般意义上的网络节点,不关注节点的个体性质;后者更强调作为网络节点的行动者并不是傀儡,而是具有"主观能动性"的行动者。我们之所以坚持这样的看法,最主要的原因在于当今国内外的社会网络学界有一种趋势,即普遍地忽视了网络节点的个体属性,特别是其能动的属性,将节点看成是受结构影响,自身不具有"反思"和"行动力"的节点,这与现实世界中的行动者是不符的。下面结合对社会网络概念的进一步说明来阐述。

(一)点:社会能动者

社会网络分析中所说的"能动者"(agent)可以是任何一个社会单位或者社会实体。能动者可以是个体、公司或者社会单位,也可以是一个教研室、学院、学校,更可以是一个村落、组织、城市、国家等。

如果能动者是"个人",可以考察他的各种"主观能动性质",包括"具体"的性质和"抽象"的性质。"具体"的性质包括诸如个人的性别、受教育年限、社会经济地位等性质。"抽象"的性质指的是个人"内在"拥有的态度、观念等抽象的属性,如个人的社会价值取向、为人之道、性格特征等。如果能动者是"民营企业",除了考察企业所在的宏观经济和治理结构之外,也可以考察它的"主观能动性质"。例如,企业领导人的慈善意识、公共意识及其带来的影响,企业的培训体系中是否体现了"人的价值";企业的管理制度是否考虑到员工的利益等。如果能动者是"国家",也可以考察它的很多"独有性质"。例如,国家的价值观(无论是坚持和平发展道路,还是扮演世界警察的角色)、国家坚守的"文化"价值等。

除了考察行动者的能动性质之外,还应该分析它所在的各类关系结构。例如,如果行动者是个人,可以考察因某种性质(如口才好、有爱心、善解人意等)带来的吸引人的程度等。这里之所以强调行动者的能动性,原因还涉及本书的性质。本书关注的是整体网,即由多个节点构成的"整体"。整体网研究对"节点"本身的关注度相对少一些,不太关心行动者的"能动性"。本小节对行动者"能动性"的阐释权且算做对本书关注点的补充。本书后文多数情况下沿用"行动者"一词,但这并不意味着我们不关注行动者的"能动"维度。"行动者"和"能动者"在本书中的含义是一样的。

（二）关系：行动者之间的联系

在社会网络研究中，为了确定关系的边界，人们常常采取两种研究路数：实在论（realist）路数和名义论（nominalist）路数。按照"实在论"的研究路数，研究者将关系视为在社会行动者之间真实存在的，对应于社会群体或组织中实际存在的关系，例如朋友之间的实际支持关系，组织内部成员的上下级关系等。按照"名义论"的策略，研究者关注的是社会行动者之间名义上的关系、感知的关系或想象的关系，所界定的关系有分析上的重要性，但它未必是被社会行动者都认识到的关系（Lauman et al.，1989；Scott，2013：42—43）。

在研究小规模社会网络的时候，人们通常遵循实在论的策略，尝试区分出一个特定群体的全部成员，追溯他们之间实际存在的各种联系。当然，如何清楚地界定所研究的关系，这里仍然存在一些问题。例如，"密友"关系对于不同的人来说意义颇为不同，因此，密友的范围可大可小。不过在笔者看来，无论实在论路数还是名义论路数，都不过是在界定关系时坚持的研究策略，据此界定的关系未必能挖掘出关系的真实"文化"内涵。例如，无论关注实际存在的社会支持关系，还是感知的支持关系（perceived social support），都不能充分地挖掘能动者在相互"社会支持"背后蕴含的丰富的"文化"要素：社会行动者到底出于什么深层次的原因而相互提供社会支持，是出于对"互惠"的承诺而提供支持，出于"公平"而提供支持，出于对弱者的"怜悯"而提供支持，还是出于利他主义倾向而提供支持，出于第三者在场而不支持他人等。我们认为，诸如此类的"文化"要素恰恰为当今社会网络研究学者所忽视，也是当今社会网络研究的软肋。

1．"关系"研究有多种。

一般来说，当我们说行动者之间存在关系（ties）的时候，"关系"常常代表具体的联络内容（relational content）或者现实发生的实质性的关系。关系有多种表现。

首先，行动者之间的关系可有一元和多元之分。"关系"可以是个人之间的朋友关系、上下级关系、喜爱关系；国家之间的国际贸易关系；城市之间的距离关系等。在一项研究中，可以关注一元关系，也可以关注行动者之间的"多元关系"。例如，两个学生之间可能同时存在同学关系、友谊关系、恋爱关系等，两个国家之间可能同时存在贸易关系、外交关系、文化往来关系等。至于具体关注哪类关系或哪些关系，完全取决于研究者的需要。

对一元及多元关系的整体网研究是当今社会网络研究的前沿领域。社会网络研究者利用多维量表（MDS）、矩阵代数（matrix algebra）、聚类分析（cluster analysis）、块模型（blockmodel）等方法来研究多元关系网络数据。也有很多学

者利用概率论、数理统计技术以及计算机技术研究网络变量的统计性质,构建多种网络模型(Carrington, Scott & Wasserman, 2005)。

其次,研究的重点不同,关注的“关系”也不同。如果研究整体网(whole network),即研究一个整体内全部行动者之间的关系,研究者就要重点分析这些关系的各种结构特征,如密度、互惠性、关系的传递性、子群结构、核心—边缘结构等。如果研究个体网(ego-network),则需要分析每个个体网成员之间关系的同质性、异质性、结构洞等多个指标。通常情况下,个体网资料可利用随机抽样方法获得,整体网资料可采用方便抽样法获取。

2. 测量“关系”的层次。

在传统的社会统计学中,数据(变量)一般分为多个测量层次:定类(两类和多类)、定序、定距、定比数据。与之类似,网络数据也分为多个测量层次。

(1) 二分类(binary)关系数据。这样的数据只有两个取值。例如,两个国家之间的建交关系只有两种取值;建交和未建交。

(2) 多分类关系数据。这样的数据取值可以多于两个。例如,两个国家之间的双边贸易额可分为多类:大、中、小。

(3) 定序关系数据。例如,两个人之间的联系紧密程度可能是:联系非常密切、比较密切、联系较少、没有联系等。

(4) 定距关系数据。例如,两个国家之间的双边贸易额数量就是定距数据。

另外有一类网络数据不能归为这种分类体系之中,这就是符号图(signed graph)数据。这种数据根据一定的标准,将行动者之间的关系赋予一定的符号。例如,在 A—B—C 三人“无向关系”结构中(如图 1.1 所示),A 和 B 之间的“+”号代表 A 和 B 相互“喜欢”;B 和 C 之间的“−”表示二者相互“讨厌”对方。在 A—B—C 三人“有向关系”结构中(如图 1.2 所示),A 指向 B 的关系符号为“+”号,B 指向 A 的关系符号也为“+”号,这表明二者相互喜欢。而 B 指向 C 的关系为“−”,C 指向 B 的关系为“+”,表示 C 喜欢 B, B 却讨厌 C。同理可分析 A 和 C 之间的关系。

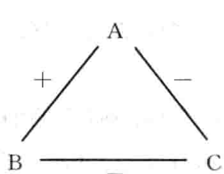


图 1.1 无向符号图

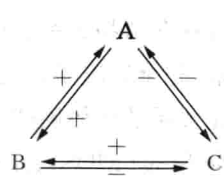


图 1.2 有向符号图

针对符号图的研究达成的共识是：符号之积为正的关系结构是稳定的结构（Heider，1958）。例如，如果A喜欢B，B讨厌C，那么C只有在也讨厌A的情况下，这三人之间的关系才是稳定的。针对符号图，可以分析其三方关系谱系、均衡性、传递性等内容。

行动者之间的关系可以是静态的、暂时的，也可以是动态的、变化的。近年来，有关网络动力学的研究逐渐增加，学者开始探讨网络变迁的动力机制（National Research Council，2003；Snijders et al.，2006，2010）。

二、1-模网和2-模网

模（mode）指行动者的集合。模数指社会行动者集合类型的数目。下面的介绍会让读者更清晰地了解"模"的含义。

1. 1-模网。

由一个行动者集合内部各个行动者之间的关系构成的网络叫做1-模网络（one-mode network）。例如，一个班级内30名同学之间的朋友关系网络就是1-模网络，世界上50个国家之间贸易往来网也是一个1-模网。

2. 2-模网。

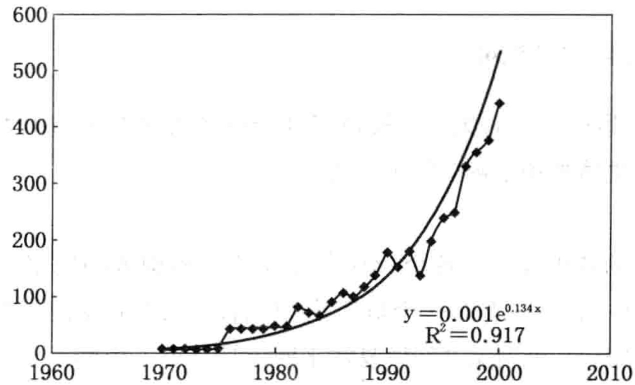
我们可能研究作为一个集合的多个公司和作为另外一个集合的多个非营利性组织之间的关系，测量资金从公司向非营利性组织的流动。网络学者把由一类行动者集合与另一类行动者集合之间关系构成的网络称为2-模网络（two-mode network）。例如，大学四年级的30名同学与大学三年级的25名同学之间的"朋友关系网"就是2-模网络，因为这里涉及两个班级的同学的朋友关系。

3. 隶属网络。

有一类特殊的2-模网络被称为"隶属网络"。具体地说，如果一个行动者集合（模态）为"各个行动者"，另外一个行动者集合（模态）为这些行动者所"隶属"的"各个部门"（如俱乐部、私人组织等），则称这样的2-模网络为"隶属网络"（affiliation network）。有关隶属网络的一个经典例子是大卫（David）和加纳（Gardner）等人于1941年对一些妇女的社会活动中存在的子群的研究，他们利用报纸、访谈记录了18位妇女参加14个事件的关系网络（参见Scott，2013：25、46）。

社会网络分析在本体论、认识论和方法论方面都有其独特之处（刘军，2004：12—26）。常规统计学处理的是属性数据，社会网络分析处理的则是关

系数据,其分析单位是“关系”。社会网络分析的核心价值在于,从“关系”角度出发研究社会行动者及其社会结构。由于社会结构既可以是行为结构、政治结构,也可以是社会结构、经济结构,所以,社会网络分析的概念和方法已经在多学科中得到了广泛应用。近 30 年来,有关社会网络方面的文章急速增加,在《社会学文摘》(*Sociological Abstracts*)中,标题或者摘要中带有“社会网络”(social network)字样的文章数目呈指数增长,如图 1.3 所示。



资料来源:Borgatti & Foster, 2003:992。

图 1.3 社会网络方面文章数的增长

社会网络除了上述术语之外,还有很多基本概念,具体参见刘军(2004:第四章)。

第二节 用图和矩阵表示社会网络

社会网络可以分为个体网和整体网两类。从分析技术角度讲,对个体网的研究已经比较完善,有关概念和理论、命题和指标等都已经比较完备。相比之下,对整体网的研究目前仍然是国外社会网络研究的重点和前沿。从数学角度上讲,有两种方法可以描述整体网:社群图法和矩阵代数方法(本部分是在刘军,2004:56—82 的基础上增删而成)。至于在表达网络数据的时候,为什么利用“正式的”图和矩阵方法,汉尼曼(Hanneman)提出了三点原因(http://faculty.ucr.edu/~hanneman/:第二章)。

首先,矩阵和图都比较紧凑简洁、有系统性(compact and systematic)。它

们都可以轻松地汇总并展示信息，这也要求我们在描述社会关系的模式方面做到系统全面。其次，矩阵和图允许我们利用计算机来分析数据。这一点十分有用，因为如果行动者数量很多或者网络维度较多的话，很难对社会网络数据进行系统的分析。大多数分析工作既耗费时间，又需要有精确性，这些工作都适用于计算机来做。最后，矩阵和图拥有自己的规则和约定。有时候，这些约定可以帮助我们进行明确的交流。更重要的是，图论语言和数学语言能够引导我们看到一些仅凭肉眼所看不到的东西。

社群图(sociogram)是莫雷诺(J.L.Moreno)在其创立的社会计量学中最早使用的。尽管由于莫雷诺的性格存在丑恶的一面(对神秘主义痴迷、言过其实的个人风格、妄自尊大)而使得社会计量学没有在20世纪30年代末受到追捧(弗里曼，2008：36—38)，但是社会计量学以及相关概念，如社群图等随后在社会网络中仍然得到了广泛使用。

本节介绍如何利用它们来表达关系网络资料。拥有了这两种工具，我们就能够理解网络分析者是怎样计算一些网络测度的(例如密度、点入度等)。

一、社会网络的图形表达

可以用图表示社会网络。图可以有多种类型，如地图、变量分布图、网络图等。与变量图不同，网络图主要由点(nodes, vertices, actors, points)(代表行动者)和线(edges, arcs, lines, ties)(代表行动者之间的关系)构成。网络图可以是多值的，也可以是二值的；可以有方向，也可以无方向；可以是1-模的，也可以是2-模的。网络分析者把根据这种思想得到的图叫做社群图(sociogram)。社群图中的点集可以表示为：$N=\{n_1, n_2, \cdots, n_g\}$。这样，一个群体成员之间的关系就可以用一个由点和线连成的图表示。

图论研究的第一步是根据现有的资料或者社会网络进行"画图"。需要指出的是，即使对同一个网络来说，用来表达它的图也可能有多种形式。例如，图1.4是表示同样关系的两个图，显然二者的形状不同。如果分析该网络的结构，可以看出二者是完全等同的。画图对于图论来说极其重要，具体的画图方法参见刘军(2004：59)。

第二步是对画出的图进行分析，可以分析图的诸多特点，如图中每个点的度数、互惠关系的数目、图的密度，分析子图、派系、三方关系分析、核心—边缘结构分析等。

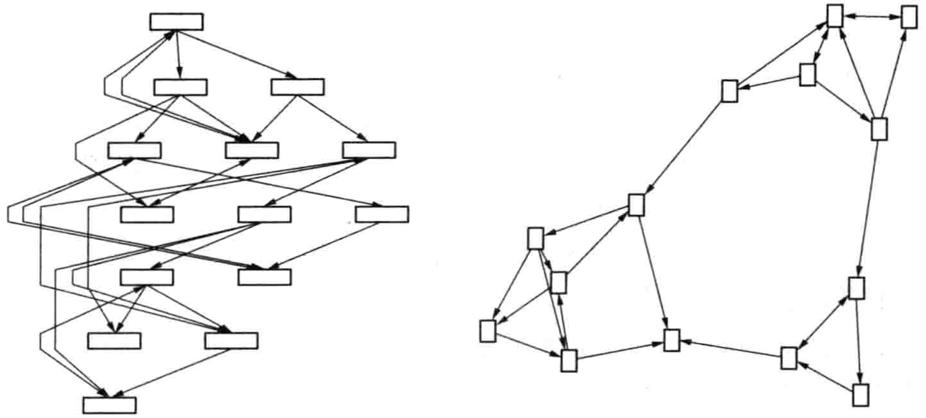


图 1.4　关系性质相同的两个图

社群图可以有多种类型，如有向图（directed graph，digraph）和无向图（undirected graph）；二值图（binary graph）、符号图（signed graph）和多值图（valued graph）；完备图（complete graph）和非完备图（non-complete graph）。完备图指的是那种其所有点都连接在一起的图。当然，这种情况即使在小网络中也很少见。也就是说，一般图都是非完备图。把关系的取值和方向结合在一起，社群图可以分为二值无向图、二值有向图、多值无向图和多值有向图等，如图 1.5 所示。

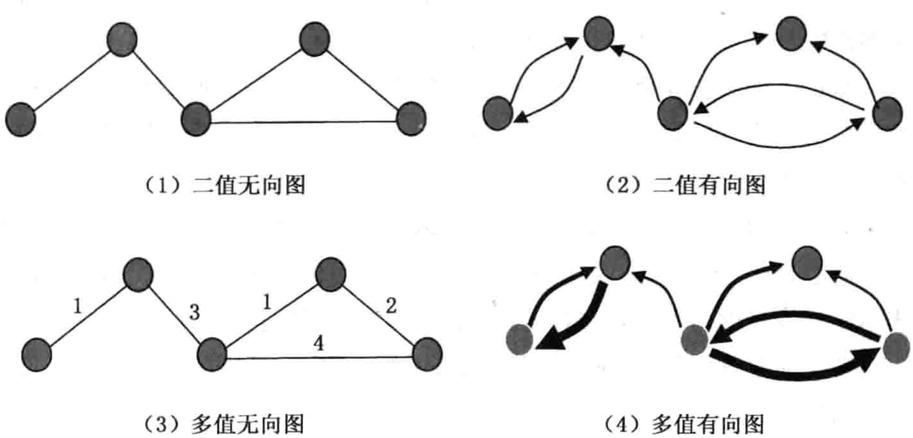


图 1.5　关系的取值与方向结合形成的四类图

根据图就可以构造矩阵。因此，与上述四类图对应，矩阵的分类也主要有四

类:二值无向矩阵、二值有向矩阵、多值无向矩阵和多值有向矩阵。

二、社会网络的矩阵表达

社群图的一个缺点是,如果社群图涉及的点很多(例如 100 人),那么图形就相当复杂,很难分析出关系的结构。在这种情况下,我们最好利用矩阵来表达关系网络。

(一) 什么是矩阵?

简单地说,矩阵就是多个因素的有规律的排列。矩阵常常用大写英文字母(例如 A)表示,矩阵中的因素用小写英文字母(如 a)表示。矩阵的规模由行和列的个数来表示。例如,一个有 3 行 5 列的矩阵 A 记作 $A_{3\times5}$。如果矩阵的行数和列数不同,称这样的矩阵为长方阵;如果矩阵的行数和列数相同,称这样的矩阵为正方阵,简称方阵。

矩阵中的要素由其所在的位置来表示。例如,矩阵 X 中的第 2 行第 4 列的要素记作 X_{24}。矩阵中的每个元素都有自己的标签或者位置。我们可以据此清楚地看到作为社会行动者的各行和各列之间的关系。

如果行和列都代表来自一个行动者集合的"社会行动者",那么矩阵中的要素代表各个行动者之间的"关系"。这种网络是前面介绍的 1-模网络。如果行和列代表来自两个行动者集合的"社会行动者",那么矩阵中的要素就代表两个行动者集合中的各个行动者之间的"关系",这种网络是 2-模网络。如果"行"代表来自一个行动者集合的"社会行动者","列"代表行动者所属的"事件",那么矩阵元素就代表行动者隶属于"事件"的情况,这种网络也是 2-模网络,具体地说是"隶属关系网络"。

(二) 社会网络分析中涉及的几类矩阵

在社会网络分析中,最常使用的是一类方阵,其中的行和列都代表完全相同的社会行动者,并且排列的顺序相同。矩阵中的要素往往是二值的。图论专家常常称这样的矩阵为邻接矩阵(adjacency matrix),并且记作 X。

除了邻接矩阵之外,还有其他很多类矩阵,如发生矩阵(incidence matrix)、隶属关系矩阵(affiliation matrix)、有向关系矩阵和多值关系矩阵等,其具体含义参见刘军(2004:63—67),此处不再赘述。

在邻接矩阵中,矩阵各个要素是"1"或者"0",分别代表关系的存在与否。根据邻接矩阵可以构建邻接关系表,如图 1.6 所示。

	A	B	C	D	E
A		1			
B	1				
C		1		1	1
D					1
E			1	1	

→

A	B		
B	A		
C	B	D	E
D	E		
E	C	D	

图 1.6　5 点图矩阵及其邻接关系

根据矩阵可以计算图的很多测度,下面仅介绍几个。

(1) 关系的总量(volume):在个体网络的意义上,一个点的关系总量指的是该点发出或者接收到的关系总数。对于有向关系来说,有点入度(指向该点的关系数)和点出度(该点发出的关系数)两种。

(2) 可达性(reachability):如果在两点之间存在一条途径(path),则称这两个点是可达的(reachable)。对于图 1.7 左侧的无向图来说,任何两个点都可达。而对于图 1.7 右侧的有向图来说,任何两个点都不是双向可达的。

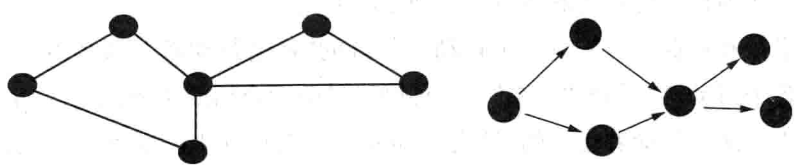


图 1.7　两个关联图

两个点之间的捷径距离(geodesic distance)是二者之间的最短途径的长度(即线数)。如果在图中任何点对(pair of nodes)之间都至少存在一条途径的话,则称这样的图为关联图,也叫做成分(component)。对于两条途径来说,如果只有两个端点是一样的,其余点不重叠的话,则称这样的途径相互独立(independent)。如果一个图在任何两点之间都存在两条独立途径的话,则称这样的图为二部关联图(biconnected),即二成分(bicomponent)。

(三) 矩阵的基本运算

在社会网络分析中,为了揭示行动者之间的关系模式,有必要对矩阵进行一定的运算和转换,主要有如下几种计算。

1. 矩阵的重排(matrix permutation),其目的是找到矩阵中隐含的有规律的关系模式。

2. 矩阵的加法和减法(add and subtraction)，即两个行和列匹配的矩阵对应元素相加减。加法可以发现行动者之间的关系的组合形式，减法可以分析关系的重叠性。

3. 一个矩阵之幂(powers of a matrix)，即一个矩阵与自身的乘积。邻接矩阵告诉我们的是在两个行动者之间存在多少条距离为 1 的途径；邻接矩阵的二次幂(即邻接矩阵的平方)告诉的是我们两个行动者之间存在多少条距离为 2 的途径；邻接矩阵的三次方告诉我们的是两个行动者之间存在多少条距离为 3 的途径，依此类推。

4. 两个矩阵之积(matrix multiplication)和布尔代数积。在社会网络分析中，矩阵的乘法是非常重要的一种运算，其含义参见刘军(2004：70—73)。可以用它来研究图中途径(walks)的个数，研究图的可达性(reachability)。

有关矩阵的各种算法在 UCINET 软件中的实现过程，请参见本书第三章的相关内容。

第三节　社会网络研究的范围和理论贡献

社会网络分析的范围广泛，内容多样，涉及社会学、政治学、经济学和管理学等多个学科。社会网络研究的范围和理论贡献可以梳理如下。

一、研究范围

社会网络研究离不开行动者及其关系的集合。不过，从目前来看，国内外社会网络学界关注的问题大都涉及社会互动的形式化表征，侧重网络结构，比较忽视网络中的行动者，这是当代社会网络分析的不足(周雪光，2003)。我们认为，社会网络研究不能脱离社会行动者、社会关系和网络结构。根据的标准不同，社会网络的研究内容也不同。

(一) 根据“关系的性质”进行分类，社会网络研究包括三类

1. 侧重关系“结构形式”的研究，关注关系的“系统”。也就是说，行动者之间的关系“模式”或“结构”怎样影响到行动者的行为，行动者反过来又如何影响结构。简单地说，这种研究把“网络”看成是变量(无论是自变量还是因变量)。此类研究很多，如：

(1) 社会凝聚力(social cohesion)的含义及测量；

(2) 关系性概念的界定,如阶级、角色、权力、自由等;

(3) 等级、控制、秩序的生成,研究哪些因素导致社会关系中的秩序;

(4) 群体内部的关系结构测量;

(5) 社会关系结构对行动者的影响;

(6) 行动者对关系及其结构的选择;

(7) 从一致性的关系活动中如何演化出角色丛(role sets)。

2. 侧重关系"内容"的研究,关注关系的"社会情境"。此类研究关注的是网络的具体"情境"如何影响行动者的行为,例子包括:

(1) 朋辈情境对犯罪有何影响? 如果一个孩子的伙伴们都抽烟,那么孩子本人是否也容易抽烟?

(2) 连锁董事和政治参与;

(3) 拥有许多弱关系的人是否更容易找到工作?

(4) 不受欢迎的孩子是否比受欢迎的孩子更容易遭遇困境?

3. 侧重关系本身"渠道效应"的研究。此类研究可以包括:

(1) "关系"通道传递的内容、性质及机制:风险、资金、资源、疾病等的传递;

(2) 创新的扩散机制:谣言、名声、新药的推广;

(3) 社会支持与精神健康;

(4) 权力的分布:核心的行动者能控制资源吗?

(二) 根据"网络类型"进行分类,社会网络研究包括三个层次

个体网(ego-network)、局域网(partial network)和整体网(whole network)。

在图 1.8 中,中心点是自我点(ego)。自我点再加上另外一个与之关联的点就构成一个"二方组"(dyad)。其他几个概念介绍如下。

1. 个体网:由一个核心个体和与之直接相连的其他个体构成的网络(如图 1.8 所示)。个体网研究的测度包括:个体网的规模(size)、关系的类型、网络的密度(density)、关系的模式(pattern of ties)、网络成员的同质性(homogeneity)、异质性(heterogeneity)等。研究的问题主要是个体网的诸多结构性质与个体的属性(如精神健康)之间有何关系。

2. 局域网:个体网加上与个体网络成员有关联的其他点构成局域网。这种网络中的关系要比个体网络中的关系多,但是比整体网中的关系数量又少。问题是,局域网的边界如何确定,这并没有定论,要看研究目的。根据不同的研究目的,可以将局域网分为 2-步局域网、3-步局域网等。2-步局域网指的是由与

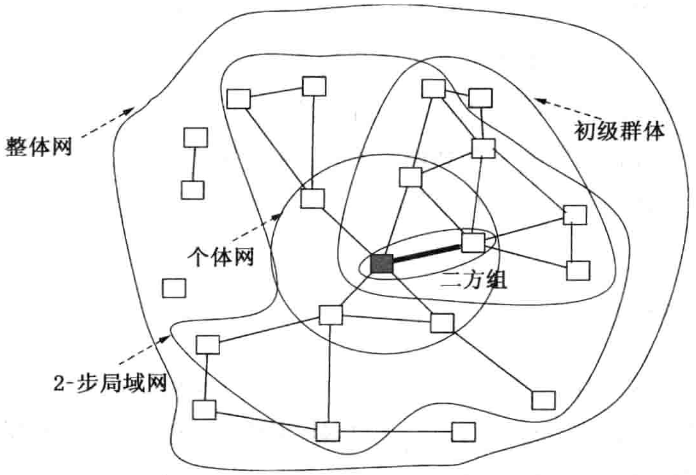


图 1.8　社会网络的多种类型

“核心个体”的距离不超过 2 的点构成的网络，该概念的含义比较宽泛，它可以包含全体到达核心点的距离为 2 的点，也可以包含部分距离核心者为 2 的点。3-步局域网指的是由与“核心个体”的距离不超过 3 的点构成的网络，4-步局域网的概念依此类推。可想而知，局域网的步长越长，网络发散得越大，这样的网络就越松散，研究起来也越难。与个体网息息相关的另外一个概念是“初级群体”，它指的是由一个人的某些关系网成员（包括直接成员和间接成员）构成的小群体。

3. 整体网：由一个群体内部所有成员及其间的关系构成的网络。整体网需要研究的测度包括：各种图论性质（graph properties）、密度（density）、子图（subgroups）、角色和位置（positions）等诸多内容，参见本书第二章。

请读者思考这样一个问题：当你在日常生活中遇到困难的时候，通常情况下会请什么人帮忙？要想回答这个问题，当然要追问具体遇到什么性质的困难。有的困难可以直接求助于朋友（例如借钱），在此意义上可以认为个体网重要；有的困难必须求助于“朋友的朋友”甚至“朋友的朋友的朋友”，在此意义上可以认为局域网重要。在当代中国，家庭已经小型化，计划生育政策导致中国人的亲属数量减少，人们遇事只能更多地依靠朋友而不是亲属。可想而知，与中国传统的大家庭相比，当代中国人在日常生活中的很多需要，特别是重要的需要是不能直接通过自己的亲属得到满足的，只能并且必须通过各类中间人的介绍和帮忙才

能得到满足。因此,在笔者看来,现实生活中重要的人际关系网往往既不是简单的个体网,也不是整体网,而是介于二者之间的局域网。换句话说,现实世界更多地表现出局域网的性质,可惜现有的社会网络研究对此并不关注。

社会网络分析的学者主要利用图论、矩阵代数、概率统计、计算机编程等方法来研究社会网络。关于图和矩阵的含义,本书有专门一章加以说明,此处不再赘述。

二、理论贡献

关于什么是"理论",这是有争议的(Abend, 2008)。社会网络分析有没有"理论"? 这一点也没有形成共识。不过,社会网络研究有一种共同的前提,即破除实体论的思维方式和研究方式,倡导关系论的思维方式和研究方式。前者将社会现象看成是实体,从作为"实体"的现象本身出发进行研究。关系论思维则考虑到关系的存在,从社会网络的角度来研究社会关系的结构,这是一种结构主义视角下的量化分析。这种关系论视角已经在政治学、经济学、社会学、心理学等诸多社会科学领域中得到了广泛应用,也在社会理论领域,诸如对权力、平等、自由、能动等方面得到了一定的体现。

当然,在解释网络动力的时候,往往也离不开文化要素的作用。例如,对分析结果的解释可能离不开社会心理学、社会学、经济学、地方性知识和规范意涵等。

另外,在笔者看来,网络分析是一种不同于因果性分析的另类研究理路,它提供了"交互"的视角(刘军,2007;刘军、杨辉,2012),倡导的不是单向因果分析,而是一种双向交互作用。这种重要的研究方法论有助于沟通微观和宏观之间的桥梁。社会网络的相关概念提供了一种构建社会结构理论的方式,这或许是社会网络分析的重要理论贡献。

社会网络研究能够为我们提供很多新的见解。例如,社会网络成员可能直接联系,也可能间接相连,而关系可能有不同的性质和内容。因此,必须在更大的网络情境中界定网络成员。社会网络一般不是随机的,而是具有一定结构,可能表现为聚类性、分派性或者六度分割、核心—边缘结构等。通过分析网络的结构,甚至可以为组织的关系管理带来直接效益。另外,通过研究可能发现,稀缺资源在不对称的关系和复杂网络中得到不同程度的分布和传递,这一洞见可能会加深我们对权力结构的理解。

本章小结

社会网络是社会行动者及他们之间关系的集合。行动者可以是个人、群体、组织乃至国家。关系是多方面的,有多种类型。由来自一个群体内行动者之间的关系构成的网络叫做1-模网络;两个群体的行动者之间的关系网叫做2-模网络;一个群体内的行动者参与多个社会团体构成的网络叫做隶属关系网。

由于关系数据不满足常规统计学意义上的"变量的独立性假设",因此,通常意义上的各种多元统计方法大多不能用来分析关系数据。社会网络分析研究的恰恰是关系数据。社会网络分析在本体论、认识论、方法论以及具体的研究方法上都有其独到之处。这也是网络分析得到学界广泛关注的重要原因之一。社会网络分析研究的内容包括个体网研究、二方关系、三方关系、凝聚子群、块模型等。研究方法包括图论和矩阵方法。

第二章

整体网研究概要

上一章已经指出，从“研究单位”角度讲，社会网络分析主要有两大研究领域：个体网研究和整体网研究。前者让我们对个体网有比较全面的认识，例如可以了解到个体网的类型、规模、同质性、异质性和构成等，探讨这些指标对个体的精神健康、幸福态度、选举倾向等有什么影响。因此，个体网研究很重要。不过，有时候我们也不能不考虑到如下事实，即诸如一个自然村落、正式组织、犯罪团伙等拥有相对明确的边界，可视其中的行动者集合为一个整体，该整体会表现出一定的结构，这些结构对其中的行动者产生影响。用社会学的术语来说，“社会结构”影响“能动作用”，这是很多社会学家，特别是坚持“结构主义”视角的社会学家所大力倡导的观点。这些问题需要利用“整体网”研究来回答。①如果说人生的幸福和精神健康与个体网有密切关系，那么与整体网的关系也不可忽视。不过，“个体网”研究一般不考察这一点。另外，当代社会网络统计技术的突破主要是在整体网领域出现的，整体网研究是国外该领域的关注点。本章主要探讨整体网研究，具体包括整体网的各类结构测度、研究方法（问卷设计、资料收集、资料分析等）和研究的意义。②

第一节　整体网研究的内容

整体网研究的内容非常丰富。例如，可以研究整体网的“互惠性”，就整体网中的“限定性互惠”和“一般化互惠”的测量方法进行研究（刘军，2007），也可以研究整体网的核心—边缘结构等。本章将有所侧重地探讨整体网的研究内容。

---

① 也可以用实验方法检验“网络结构”对“个体行为”的影响，这是一种社会心理学的实验研究，与严格意义上的整体网研究有所不同（参见刘军、郭丽娜，2013a，2013b）。

② 关于整体网研究，特别是整体网资料的收集方法，请读者参见罗家德（2010：63—89）。本章有少许内容参考了该书，当然也有所补充。

一、整体网的分类

如果根据社会网络中的"行动者"类别进行分类，整体网也有很多类型。例如，如果行动者是个人，对应的整体网就相当于很多"个人"之间的"人际关系网"，具体如"朋友关系网"等；如果行动者是组织或者法人行动者，对应的整体网就是"组织关系网"；如果行动者是城市、村落、国家等，对应的整体网就是"城市网"、"村落网"和"国际关系网"等。正是在此意义上，我们说整体网有多种类型。我们这里主要结合 1-模关系探讨整体网研究的内容。

就行动者之间关系的具体内涵来说，整体网又可分为各种类型。如就"国际关系"的具体内容来说，可以把国际关系网分为"国际贸易网"、"国际科技往来网"、"国际政治关系网"等；就"企业关系网"来说，可以细分为企业之间的"营销同盟网"、"财务同盟网"、"战略联盟网"等，一个企业内部的各个员工之间还存在"情感关系网"、"咨询关系网"、"情报关系网"和"信任关系网"等(罗家德，2005：64)；一个村落内部各个家庭之间存在"情感支持网"、"资金支持网"、"劳力支持网"和"亲属关系网"等(刘军，2006a)；多个学术期刊之间存在"相互引用关系网"。另外，由多个行动者参与多个团体构成的网络，即"隶属关系网"也可以看成是整体网。

整体网研究特别是 1-模网络研究已经取得很大进展，近 30 多年在统计方法上有重大突破，例如 p^* 模型、动态网络模型等，这些研究加深了对整体网结构的认识。

二、整体网的构成和规模

(一) 整体网的构成

在整体网中，也要研究网络成员的构成，研究方法类似于对个体网构成的研究。例如，我们可以分析在整体网中的每个行动者的个体网络成员由哪些类型的行动者构成，比如性别结构、年龄的同质性、受教育程度的异质性。此处不再赘述。

除此之外，我们还可以按照某种关系来分析另一类整体网的构成。例如，在研究"公司内部成员之间的建议关系网"的时候，研究者可以根据自己感兴趣的项目对网络的构成进行分析。如果根据"上下级关系"来分类，可以分析在"建议关系网"中，存在多少"同级别职员之间的建议关系"，"不同级别职员之间的建议关系"等。实际上，这种分析类似于在"二方组"研究中所分析的那样，研究整个

网络是由多少个"互惠对"、"不对称对"和"虚无对"构成的。我们也可以研究三方组的16个同构类在整个网络中有哪些表现,这也是对整体网构成的研究(参见刘军,2006b)。实际上,关于整体网的各种结构的研究都属于网络构成研究的范畴,我们将在本节第五部分,即关于整体网的结构研究部分加以介绍。显然,这种研究完全不同于个体网的构成研究。

(二)整体网的规模

个体网的规模指的是与某个核心个体直接相关的其他个体的数量,往往需要调查才能有所了解。整体网的规模指的是网络中包含的全部行动者的数目。例如,如果研究一个企业内部各个员工之间的"建议关系网",那么该企业内部的全体员工总数就是该整体网的规模。如果研究50个国家之间的"贸易关系",该网络的规模就是50。可见,整体网规模的含义不同于个体网规模的概念,在计算个体网规模的时候,一般不包括作为核心个体的"自我"(具体参见刘军,2004:102—104)。

整体网的规模越大,其结构通常也越复杂,分派现象越普遍,派系、n-派系、k-丛等也往往越多,研究起来也比较困难。在当代的整体网研究中,一般情况下网络规模不会超过1 000。实际上,现实的社会世界很小。有时候研究大量(如超过1 000)行动者之间的整体网也没有什么实际意义。例如,研究一个城市社区内部几千人之间的关系网络,这种研究一般没有太大的学术意义,也没有什么实践价值。研究者往往对具有较大社会学意义的相对封闭的整体进行研究,而这种整体网的规模一般不大。

就人际关系来说,个体与外界可以保持稳定的社会关系,这里所说的"稳定"可以理解为群体之人相互认识并知道其他人之间的关系。可想而知,个体不可能与非常多的"他人"维持稳定的社会关系。英国牛津大学人类学教授罗宾·邓巴(Dunbar, 1992)指出,人类的社会世界是很小的世界。5人左右为亲密接触圈,12至15人为同情圈,50人左右的群落是经常一起生活、行动的人,150人左右的规模是遵从共同仪式的人。邓巴认为,一个社会群体合适的规模大致是150人。也就是说,有效地记忆、管理、组织的群体规模应在150人左右。超过这个数字,一个群体成员之间的互动就会迅速减少,无法有效地沟通或协作,需要更多法律、政策、规则去维持群体内稳定的凝聚力。150这个数字就是著名的邓巴数(Dunbar's number)。当人数上升到500人的时候,需要有同种语言才可沟通,上升到5 000人时,维系社会结构只能依靠共同的文化。邓巴数既可以表示适当的个体网规模,也可以表示有实际意义的整体网规模。

三、整体网的密度

关于整体网密度的计算，笔者已经有所论述（参见刘军，2004：第 4 章），这里将有所补充，并且结合 UCINET 软件，具体展示计算整体网密度的步骤。

整体网密度的计算与个体网密度的计算不同。如果整体网是无向关系网，其中有 n 个行动者，那么其中包含的关系总数在理论上的最大可能值是 $n(n-1)/2$。如果该网络中包含的实际关系数目为 m，那么该网络的密度就是"实际关系数"除以"理论上的最大关系数"，即等于 $m/(n(n-1)/2)=2m/(n(n-1))$。如果规模为 n 个行动者的整体网是有向关系网（directed graph），那么其中包含的关系总数在理论上的最大可能值是 $n(n-1)$。该网络的密度因而等于 $m/(n(n-1))$。

总体而言，整体网的密度越大，表明网络成员之间的联系越紧密，该网络对其中行动者的态度、行为等产生的影响就越大。联系紧密的整体网不仅为其中的个体提供各种社会资源，同时也成为限制其发展的重要力量。关于这一点，当代整体网研究技术已经能够检验此类命题的统计显著性（参见 Wasserman & Faust，1994；刘军，2006a；Lusher，Koskinen & Robbins，2012）。

利用 UCINET 软件计算网络密度，可分别计算"个体网"的密度①和"整体网"的密度，二者含义不同。

（一）整体网中的"个体网密度"的计算

如果计算整体网中每个点的个体网密度，需要在 UCINET 中沿着 Network→Ego networks→Egonet basic measures 这条路径，然后选择所要分析的网络数据，即可计算出一系列个体网的测度指标来，包括网络密度。

（二）整体网密度的计算

二值关系网密度的计算公式比较简单，它等于网络中"实际存在的关系总数"除以"理论上可能存在的最多关系总数"。UCINET 还可以计算多值关系网络的密度，其计算方法与二值关系网密度的计算公式一样。另外，该软件也可以计算出不对称矩阵的密度。

---

① 在一般的个体网研究中，每个个体网的"核心点"都是根据随机原则抽取出来的。这里所说的整体网络中的"个体网"是从整体网中单独抽出"个体"，考察该个体的网络成员得到的网络，它不是随机的，因而不同于通过随机抽样方法得到的个体网。当然，这两种个体网密度的计算公式相同。

在 UCINET 中,依次点击 Network→Cohesion→Density→Density Overall 这条路径,然后在出现的对话框中选定要分析的网络数据(如"紧密关系表"数据),即可计算出该网络的整体的密度。

四、整体网成员之间的距离

在整体网中,两点之间的距离是图论意义上的距离,即二者之间在图论或者矩阵意义上最短途径(即捷径)的长度,不同于个体网研究中的"距离",也不同于当代社会中人与人之间越来越大的心理距离概念。在计算一个整体网成员之间的距离时,需要计算出整体网矩阵中所有点之间的"距离矩阵"或者"一般化的距离矩阵"(a distance or generalized distance matrix between all nodes of a graph),同时也允许把"距离矩阵"(其中的值越大表示关系越疏远)转变为"接近矩阵"(其中的值越小表示关系越疏远)。所谓两个点之间的"一般化的距离"(generalized distance)指的是这两个点之间存在的一个最优途径的长度(the length of an optimum path)。两个点之间可能存在多条途径,哪条途径是最优途径?这里存在如下三种标准:"费用最小"、"强度最大"和"概率最大"(参见 UCINET 中的"help")。

1. 费用最小的途径是最优途径。一个途径包含多条边,每一条边都有其取值,即边值。途径的"费用"(cost)是该途径上的所有边值之和。在两个点之间存在的多条途径中,费用最小者为最优途径。

2. 强度最大者为最优途径。一条途径的"强度"或"能力"(strengths/capacities)指的是该途径中各条线的强度或者费用。最优的方法是采用强度(或费用)最强的途径(the optimum is the strongest path)。

3. 一条途径的概率是其各条边的概率(probability)之积,最优途径是"最可能出现的途径"(the most probable path)。

如果在两个点之间存在的最优途径多于一个,那么 UCINET 算法会采用最短的最优途径(the shortest optimum path)。对于二值邻接矩阵来说,"距离"和"一般化的距离"是等同的。在实际分析时,UCINET 程序要求研究者自己给出"距离"的类型,默认的类型是"邻接距离",即对应于图论距离意义上的标准化二值数据。除此之外,可供选择的类型有上述"Strengths"、"Costs"、"Probabilities"等三类。

另外,"距离矩阵"也可以通过一种"接近度转换"步骤(a nearness transformation)被转换为"接近矩阵"(nearness matrix)。软件系统本身的默认值是"不分析",即不

进行"接近度转换"。如果希望转换，可采用的方法包括取倒数、线性转换、指数化等。可供选择的项目有："Multiplicative"，即用最大可能的距离($N-1$)除以两个点之间的距离，得到的新值为$Y_{ij}=(N-1)/D_{ij}$。"Additive"，即"加减运算"，用点总数N减去两个点之间的距离，得到的新值为$Y_{ij}=N-D_{ij}$。"Linear"，即两个点之间的距离线性地转换为[0，1]之间的数字，得到的新值为$Y_{ij}=1-(D_{ij}-1)/(N-1)$。"Exponential"，即两个点之间的距离根据指数递减性(exponential decay)进行转换，得到的新值为$Y_{ij}=bD_{ij}$。衰减因素b(the attenuating factor b)需要由研究者自己选定，并且需要满足$0<b<1$这个条件。"Freq Decay"是利用社会学家伯特(Burt，1976)于1976年提出来的频次衰减函数(frequency decay function)进行转换。如果选择了它，那么"Attenuation Factor"的默认值是0.5。该值越大，衰减速度越慢。

需要注意的是，两点之间的距离指的是二者至少可通过多少条边关联在一起。因此，在整体网中，计算出的所有行动者之间的距离要用"距离矩阵"(distance matrix)表示。例如，假设有一个由五个行动者组成的网络，命名为5nodes，保存在D盘的datafiles文件夹里。具体的网络关系取值如表2.1所示。

表2.1　5点图的社群矩阵

0	1	0	1	0
1	0	1	1	0
0	1	0	1	0
1	1	1	0	1
0	0	0	1	0

如果计算这5个点两两之间的"邻接"距离，需要点击Network→Cohesion→Geodesic Distance，在出现的对话框中"Input dataset"项选出D盘文件夹datafiles中的5nodes，如图2.1所示：

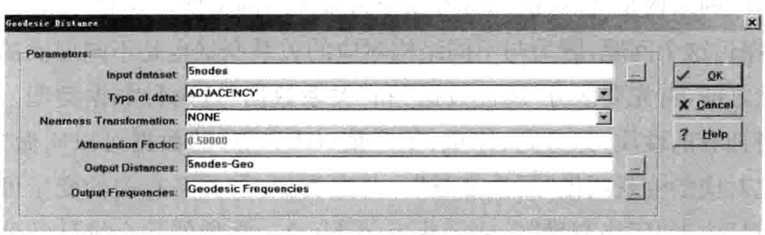


图2.1　计算距离的对话框

点击OK之后，得到如下计算结果：

表 2.2 距离计算结果

```
Average distance                                  = 1.400
Distance-based cohesion ("Compactness")           = 0.800
 (range 0 to 1; larger values indicate greater cohesiveness)
Distance-weighted fragmentation ("Breadth") = 0.200

Frequencies of Geodesic Distances

        1  2
       Fr Pr
       -- --
  1    12  1
  2     8  0

Geodesic Distances

        1 2 3 4 5
        - - - - -
  1     0 1 2 1 2
  2     1 0 1 1 2
  3     2 1 0 1 2
  4     1 1 1 0 1
  5     2 2 2 1 0
```

可见,该程序计算出任何两个点之间的距离,得到的平均距离为 1.40,建立在“距离”基础上的凝聚力指数为 0.80。该指数越大,表明该整体网越具有凝聚力。因此,可以认为这五个点之间的关系比较紧密,凝聚力强。

五、整体网的结构研究

“结构”一词的含义非常广泛。广义地说,本节第二部分已经探讨了整体网的构成,这可以说是从“纵向”对网络的结构进行分析。对三方关系、凝聚子群的研究等也都是一种整体网的结构研究。我们认为,如果深入到整体网的每一种“构成”,从“横向”进行分析,探讨关系网络的内部构成,得到的将是对整体网“深层次结构”的认识。这种研究是个体网研究所做不到的。

整体网的结构是“深层次”的,可以分为多种结构类型。我们可以结合“关系变量”来探讨这个问题,因为与不同结构对应的关系变量也是不同层次的。在进一步讨论之前,首先要区分“属性数据”和“关系数据”这两种数据类型。关于行动者自身拥有的性质(如身高、体重、年产值、GDP 等)的数据可以叫做“属性数据”,与之对应的变量叫做“属性变量”。与之类似,与关系有关的变量可以称为“关系变量”或者“网络变量”(network variables)。常规的社会统计学分析的大都是属性数据和属性变量,而社会网络分析则尤其适用于分析关系数据及网络变量(参见刘军,2004, 2006a)。

（一）属性变量和网络变量

在常规的社会学量化研究中，我们遇到的变量如"性别"、"年龄"、"收入"、"GDP"、"师生比"等都可以看成是该行动者（不管该行动者是个人、企业、国家，还是大学）本身独有的变量，可称之为"属性变量"（参见刘军，2004，2006a），相应的分析单位（如个人、企业、国家等）是相对独立的"点"。与之对应，网络变量指的是与整体网络的结构性质对应的变量，它至少可以包括如下五类：(1)个体层次网络变量；(2)二方关系网络变量；(3)三方关系网络变量；(4)子群层次网络变量；(5)整体网络变量。

通过分析"属性变量"，可以揭示行动者的属性数据之间的关系，而通过分析"网络变量"，则可以揭示行动者之间的社会结构。对"关系数据"或网络变量进行整体结构分析很重要，这一方面是因为"关系"既影响网络成员的行为，也影响行动者系统的行为；另一方面，"关系数据"不能还原到属性数据。例如，关系好的人之间可以提供帮助，凝聚力高的群体会影响成员的行为，这是仅仅利用属性资料所不能解释的。关系存在于至少两个行动者之间，因此，"关系"不是单个行动者拥有的属性，而是至少两个行动者共同拥有的属性。就此而言，我们说关系数据不能还原到属性数据，"关系"是一类独特的分析单位，不同于常规的分析单位。对属性变量和网络变量、"点"和"关系"的这种分析如表 2.3 所示。

表 2.3　两类分析单位及其研究内容对照表

单位	举　例	研究内容①	使用的变量举例	量化方法
点	个体、群体、组织等	"点"的属性特征，如年龄、声望、GDP 等	"属性"变量，如收入、人口数、GDP 等	常规统计技术
关系	个体之间、社区之间……的关系（1-模关系）；两组行动者之间的关系等（2-模关系或"隶属关系"）	网络特征，包括： 1. 个体层次：接触频次、异质性、扩张性等； 2. 双边层次：互惠性等； 3. 三边层次：关系的传递性等； 4. 子群层次：子群内的互惠性等； 5. 整体层次：关系总量等； 6. "关系"的演化	网络变量，可以包括： 1. 个体属性变量； 2. 个体网络变量：聚敛度、扩张度等； 3. 二方变量：互惠度等； 4. 三方变量：传递度等； 5. "块"网络变量：如块内互惠度等； 6. 整体变量：总选择量等； 7. 时间变量	社会网络分析技术，包括： 1. 中心性分析； 2. 核心—边缘分析； 3. 块模型； 4. p^1 模型； 5. p^* 模型等； 6. 动态网

① 此处所说的社会网络的研究内容主要指的是"整体网"的研究内容。"个体网"研究主要关注的是个体的各种网络属性特征，如个体网的规模、个体网的构成等。

(二) 整体网络结构研究的内容

整体结构研究除了上述内容之外,还有很多。这些研究或多或少都与各层网络变量相对应,整体网络结构研究因而可以具体分为至少如下几种类型。

1. “中心性”分析——权力的量化研究。

在什么意义上说一个“行动者”有权力,在什么意义上说一个“子群体”拥有权力,在社会网络中是通过研究“点”或者“群体”的“中心度”(centrality)和网络的“中心势”(centralization)指数来探讨这个问题的(Freeman, Borgatti & White, 1991)。这种研究具体包括:度数中心度和度数中心势;中间中心度和中间中心势;接近中心度和接近中心势;特征向量中心性研究和权力指数。另外,“中心性”研究又有进展(Bonacich & Lloyd, 2001; Costenbader & Valente, 2003),包括群体中心度的研究(Everett & Borgatti, 1999; Bonacich et al., 2004)、个体中心度和群体中心度的比较(Marsden, 2002)、2-模网的中心性研究(Faust, 1997)等。另外,利用实际网络数据展示各种中心性方法的应用,也是中心性研究的实证研究内容,这种研究也具有一定的现实意义。对“中心性”的研究可以加深对权力的量化认识(参见本书第五章)。

需要指出的是,中心性研究是根据网络数据计算中心度和中心势,是基于关系数据计算出来的,是从数据中挖掘出来的。这种计算思路及其结果与现实世界中某些行动者可能因处于高“地位”而居于“核心位置”有所不同。

2. 二方关系研究。

社会网络意义上的二方关系研究主要是定量研究。但是笔者认为,对二方关系也可以进行定性研究(参见刘军,2006a)。就定量研究来说,二方关系要研究的问题有:在一个社会网络中,“互惠关系”出现的频次多大,在统计上是否显著?“虚无关系”、“不对称关系”等出现的频次是多少?在统计上是否显著?所有这种研究都与“二方谱”(dyad census)有关系。就定性研究来讲,二方关系需要研究的问题可能有:二方关系有哪些表现(竞争、合作,还是互不干扰)、二方关系的演变过程及机制是什么?结合具体的二方关系(如夫妻关系、朋友关系、上下级关系、两个国家之间的关系等),二方关系表现出怎样的特点?这些问题既可以进行定量分析,也可以进行定性探讨。

3. 三方关系研究。

与二方关系研究类似,对三方关系研究也可以进行定量研究和定性研究。就定量研究来说,需要探讨的问题主要有:在一个社会网络中,三方关系有哪些

表现？三方关系的 16 个同构类(即三方谱)分别发生多少次？在统计上是否显著？另外，对某种三方关系命题进行统计检验也属于定量研究。

就定性研究来说，三方关系要研究的问题有：三方关系中的行动者如何“拉关系”，维持关系，如何形成“二打一”(two against one)的局面？根据社会行动者权力分布的不同，三方关系有哪些表现，三方关系未来的走向是什么，三方关系中如何出现冲突和合作，关系对网络联盟有什么影响等。有关这方面的研究参见网络交换论的研究成果(Willer，1999；刘军、郭丽娜，2013a，2013b)。

4. 小群体的量化研究——块模型分析。

块模型(blockmodel)最早由怀特等人(White，Boorman & Breiger，1976)引入社会网络分析之中。该模型关注总体的网络结构，在理论和实践上都具有重要意义。在块模型研究中涉及的内容有：如何界定块模型，构建块模型的两种方法：Concor 以及层次聚类方法，如何解释块模型分析的结果(Wasserman & Faust，1994：397—401)等。另外，随机块模型(Wasserman & Anderson，1987)和推广的块模型(Doreian，Batagelj & Ferligoj，2000；Doreian，Batagelj & Ferligoj，2004；Doreian，Batagelj & Ferligoj，2005)近些年得到研究。块模型有广泛的应用性，如应用该模型来分析一些现实数据，或者分析一个村落的社会支持数据、国际贸易关系数据等。

块模型研究的目的是找到在总体网络中存在的子群体。由于网络数据类型不同，结构多样，因此，这种研究也分为多类。

5. 凝聚子群分析。

“社会团结”、“社会凝聚力”是社会学研究的经典问题之一。如何定量地研究“社会团结”？我们可以通过研究各种凝聚子群探讨这个问题(Hummon & Doreian，2003)，我们可以结合具体实例进行应用研究。凝聚子群研究的内容包括：凝聚子群研究的渊源；凝聚子群的几种类型(派系、n-派系、n-宗派、k-丛、k-核等)，这些子群之间的关系；复杂网络中的各种凝聚子群研究；利用 UCINET 中的凝聚子群算法，分析现实网络数据(如一个国有企业中层领导之间的咨询网络)中存在的凝聚子群。这种研究有助于揭示现实生活中存在的“人浮于事”、“效率低下”等现象的原因，并提出对策。

6. 多个行动者之间的对等性分析。

“位置”和“角色”是“关系性”概念。通过对关系数据的分析来探讨行动者的位置和角色，这或许是社会网络分析对社会学理论最重要的贡献之一。通

过考察社会网中的社会行动者所处的“位置”和“角色”,可以分析到个体在社会网中的重要程度。这种研究包括:“结构对等性”(structural equivalence)分析;“自同构对等性”(automorphic equivalence)分析和“规则对等性”(regular equivalence)分析(Wasserman & Faust, 1994; Boyd & Jonas, 2001)。

结构对等性研究的目的是找到在结构上对等的点集合,从而对整体网有更详细的认识,也有助于对日常生活中常见的“角色”这个概念进行深入的量化认识。

7. “核心—半边缘—边缘”结构分析。

关于“核心—边缘”的思想散见于各种文献之中。现代世界体系理论提出的“中心—半边缘—边缘”结构是对西方经济发展理论和依附理论的“中心—边缘”结构的继承,但是该理论没有给出分析这种结构的具体方法,这恰好是社会网络分析中“核心—边缘(core-periphery)结构研究”的贡献。该模型可以揭示在一个社会网络中哪些行动者处于“核心”、“半边缘”和“边缘”地位,并找到这些结构之间的关系。

“中心—半边缘—边缘”结构分析的研究内容包括:“核心—边缘”模型的构建,艾弗雷特等学者在此基础上引入“核心—半边缘—边缘”模型(Borgatti & Everett, 1999)。我们可以利用这种模型进行实证研究。例如,可以结合一定的现实例子(如中国与世界其他国家之间在21世纪初的科技交往、文化往来、贸易往来资料),利用UCINET软件分析国际贸易的“核心—边缘”结构,分析21世纪的中国在国际关系中的具体位置(参见刘军,2004)。当然,这需要收集大量的资料。又如,我们可以分析一个学院、组织内部的核心—边缘结构,从而为有关政策的制定和实施提供参考。

8. 网络数据的统计推断——指数随机图模型(p^*)研究。

指数随机图模型(exponential random graph models)(简称p^*模型)是社会网络模型研究的最前沿问题之一,其初始模型是p^1模型,近30年又发展出p^*模型及各种推广形式。这种模型研究的目的是把各种网络变量和属性变量都加入到模型之中,分析哪些网络变量或属性变量对于“关系”的发生有显著影响。

p模型目前的研究内容包括:基础模型,即p^1模型及其推广形式的含义;p^1模型的推广,即p^*模型;p^*模型的扩展,包括隶属网络的logit模型研究(Skvoretz & Faust, 1999; PengWang et al., 2009)、多值网络的p^*模型分析

(Robin, Pattison & Wasserman, 1999; Robin, Pattison & Woolcock, 2004)。另外,社会影响模型(Robin, Elliot & Pattison, 2001)研究哪些网络变量影响到关系的发生,而社会选择模型(Robin, Pattison & Elliot, 2001)研究哪些变量影响人们做出的选择。可以利用推广的 p^* 模型分析现实数据,如利用该模型分析市民或者村民的社会支持网络数据,建构社会支持网络模型。

9. 偏好网络论研究。

有关偏好网络论(biased net theory)的研究已经有几十年历史,是社会网络模型研究最早建立的模型之一。它所针对的网络主要是大型的网络。这种研究的目的是:分析网络结构的变化在多大程度上是由行动者之间非随机的偏好,即"关系事件"带来的,又在多大程度上是"随机因素"造成的。所谓偏好在这里是指行动者的行为倾向。例如,如果人们坚持认为"人不犯我,我不犯人;人若犯我,我必犯人"、"己所不欲,勿施于人"或者"来而不往非礼也"等行为规范,那么这些相当于行动者之间的"互惠"偏好。如果认为,朋友的朋友就是朋友,这就是关系的"传递性"偏好。因此,可以研究一个行动者的行为在多大程度上受到某些偏好如"互惠性"、"传递性"、"同质性"等的影响,行为的随机性又有多大?利用这种研究可以对有关命题进行检验和统计分析。

偏好网络理论最先是由拉波波特(A.Rapoport)于 20 世纪 50 年代提出来的,后来由法拉罗以及斯克弗雷兹等学者重新提出(Fararo, 1981, 1983; Fararo & Skvoretz, 1984; Skvoretz, 1985, 1990)。偏好网络模型的构建需要考虑到两类参数(结构参数和属性参数),需要对这两类参数进行估计(Skvoretz, Fararo & Agneessens, 2004)。这种研究需要结合实际数据加以分析。这也是整体网研究的内容之一。

10. 网络演化研究。

上述研究都是"静态"的研究,这种研究虽然可以揭示关系网络的结构,但是不能揭示现实关系数据的动态演变。因此,从"动态"的角度出发,考察至少两个时段之间的行动者网络结构的变化(参见 Snijders, 2003; Snijders et al., 2010),这种研究可以分析出有哪些网络因素或属性因素带来社会关系的变化。

整体网研究还有相当多的内容,如社会影响研究、社会选择研究、复杂网络研究、社会均衡研究、交换网络研究等,限于篇幅,这里不一一做介绍。

第二节　整体网资料的收集

与“个体网”研究类似,在进行“整体网”研究之初,也要考虑方法论和具体的研究方法问题。例如,样本怎样选取,问卷怎样设计,数据怎样整理和分析等,下面将探讨这些问题。首先需要指出的是,1-模网络数据是最常见的,易于收集。而1-模局域网络数据则很少见,因为调查的成本高。下面的分析主要针对的是1-模网络数据。

在进行社会学量化研究时,资料收集的方法多种多样,可以利用抽样调查法、典型调查法、方便抽样法、实验法,也可以利用文本、历史资料、互联网资料等。这些方法也可用来收集网络资料。

一、整体网资料的收集方法

(一)利用“线人”收集资料

研究者通常根据自己的兴趣点(而不是随机选择)选取要研究的整体。因此,整体网资料的收集不利用概率抽样的方法,而是利用线人(informants)方法收集资料,即研究者通过与多个线人进行长时间的访谈,记录被研究对象之间的各种关系,构建关系矩阵并进行分析(具体参见刘军,2006a, 2006b)。

在田野研究中,如果研究的目的是理解当地的社会情境(例如了解某个帮派或者团伙),研究者就需要与当地人进行长时间的接触和深入的访谈。线人就是那些可以为研究者直接提供有关他所在群体的丰富信息的人,这些人首先应该是该群体的成员。在人类学研究中,利用线人收集资料这种方法尤其重要。如果研究者希望研究一个地方社区内部的非政治关系网,那么利用线人的方法无疑是可采用的重要方法之一。因为线人对他所在的社区有深入的了解,能够提供比较深入的信息,如哪些家庭之间的关系比较好或不好,哪些家庭发生了重大事件等。这些信息往往是通过问卷调查所得不到的。

(二)根据提名法收集资料

整体网资料可以根据访谈法或结构式问卷法加以收集。与个体网问卷的设计类似,在设计整体网问卷时,可以包含两个部分:提名生成法(name generator)和提名诠释法(name interpreter)(参见罗家德,2005:46—47)。

笔者认为,提名生成法不仅适用于收集个体网资料,也可用来收集整体网资

料。因为如罗家德(2005:46)所说,提名生成法是请被调查者说出与自己有某种关系的一些人,再说出这些人之间是什么关系。但是,在整体网资料的收集过程中,可采用两种策略:回忆法和名册法。所谓回忆法(recall survey)指的是邀请被调查者回忆与自己有关系的人有哪些。但是,考虑到被调查者可能遗忘一些人,因此,在调查之前可以准备好该整体内部的所有成员,形成名册(check-list format),然后邀请被调查者根据名册指出与自己有关系的人有哪些。但是,需要指出的是,不管研究者如何努力,一般都很难收集到完备的资料。

可以看出,利用这种方法得到的数据是整体网数据。由于并非所有的行动者都对自己关系网络成员之间的关系一清二楚;又由于多个行动者对 A、B 两个人之间关系好坏的判断不一定一致,在这种情况下如何分析数据的好坏呢?我们固然可以采用多种方法进行加权。但在笔者看来,"线人"方法可用来补充收集更详尽的信息,因为线人是消息灵通人士,对人际关系了解得比较清楚。

当然,整体网研究中的"提名生成法"与个体网研究中的"提名生成法"还是有区别的。就前者来说,被调查者往往在其所在的整体内部选择与自己有关系的行动者,而不像在个体网研究中那样,可以随意选取;就后者来讲,被调查者首先选择其个体网成员,再根据自己的经验判断这些人之间是什么关系。显然,就后者来说,每个"自我"是不同的,每个"自我"选择出来的"个体网成员"可能属于完全不同的群体,因而可能根本不认识,也就无法构成一个整体网研究的对象。所以,这两类提名法是不同的。

需要补充的是,有学者注意到,在利用提名法收集网络资料的时候,被调查者往往倾向于说出"积极性的关系",例如愿意说出哪些人帮助过自己,不愿意说出"消极性的关系",例如不融洽的、感觉困难或冲突性的关系就是被调查者不愿意回答的(Marsden, 1990:442)。这一点对于社会支持网络研究来说尤其要注意。

(三) 利用职位生成法收集整体网资料

个体网数据的收集可采用"职位生成法",这种方法也可以用来收集整体网资料,不过学者们利用的比较少。可以想象,如果研究者希望研究整体网中社会资本的分布情况,探讨社会资本与网络特征之间的关系,那么职位生成法是可以应用的,但是要做相应的调整。在个体网研究中,利用职位生成法需要询问被访者在诸多职位中有多少接触者。在整体网研究中如果利用职位生成法,需要询问的问题是:在整个网络中,与被访者接触的人都是什么职位的人,当然在研究

之前要先对各个职位的具体含义进行界定。

利用职位生成法可以研究的问题有：不同职位的行动者与哪些职位的人有接触？或者说，行动者的权力与接触者的权力是什么关系？对此问题可以这样研究：在一个整体网络中，首先分析与行动者接触之人的职位分别是什么，然后对各种职位的等级进行"排序"，计算每个个体所接触之人的职位同质性或异质性，根据"职位等级"计算整个网络的"密度"，计算网络关系在职位等级之间的传递程度等。例如"科学家"的职业声望最高，"教授"职位次之，"律师"相对较低一些。因此，可以按照职业声望的高低，对行动者所接触的人进行排序。其方法如下所示。假设有一个由六个行动者 A、B、C、D、E、F 构成的整体网，他们的职业都已知(假设 A 是科学家，职业声望值是 95；B 是教授，声望值是 90；C 是企业家，声望值是 85；D 也是科学家，声望值是 96；E 是律师，声望值是 78；F 是工程师，声望值为 80)，因此排序也已知。因此，由这六个行动者构成的整体网的元素 x_{ij} 可按照如下方法得到：$x_{ij}=(i$ 的声望值$)/(j$ 的声望值$)$。当然，从这种分析中可以看到，$x_{ij}\cdot x_{ji}=1$。即关于主对角线对称的矩阵元素之积等于 1，这是由我们关于该矩阵的定义所决定的。

构建了这种矩阵数据之后，我们可以分析其中的各种结构特征，如可以分析上文介绍整体网研究中的各种特征；也可以分析"声望矩阵"和"接触频次"矩阵之间的相关系数，这就需要利用 QAP 方法(参见本书第十二章，即"关系数据的假设检验"一章)。

需要指出的是，利用这种收集资料的方法进行的整体网研究成果并不多。因为，一般来说，整体网络往往针对一种关系(如只针对出口关系、资金支持关系等)进行分析。

(四) 利用档案资料收集整体网数据

利用调查法或问卷法收集的资料(包括网络资料)往往不完善，并且费用高，而利用档案资料收集整体网数据的做法则比较省钱，因而在国外已经得到应用(Padgett & Ansell, 1993; Gil-Mendieta & Schmidt, 1996; Gould, 1991)，但是在国内的应用很少。利用档案可以收集到过去发生的、其他方法收集不到的关系资料。

例如，有学者利用一些传记词典(如《美国女名人(1607—1950)》、《世纪女杰》、《名人录(1914—1915)》)研究了 19 世纪美国女权运动中各个组织之间的关系(Rosenthal et al., 1985)。作者考察了女权改革的领袖所隶属的组织情况，建

构各个组织之间的关系矩阵，测量矩阵的密度和方向性，分析哪些组织之间有相近性，哪些组织居于网络的核心，从而找到了多个聚类。该研究清晰地展示了网络分析技术可用于研究历史数据，且有助于深入理解社会运动。发现1840年至1914年的女性参与的诸多改革运动之间是高度关联的，各个组织之间的关系主要是弱关系，当去掉最核心的三个组织以及少于三个的关系之后，整个网络便会出现不同的聚类。

在科学社会学中，有学者研究引用关系，进而确定科学共同体的具体情况。如果两个科学家的文献被同样的作者引用的话，我们就说他们之间存在着关系。关于引文关系的研究可以揭示哪些作者居于核心位置，哪些作者居于边缘位置。此方面研究的经典文献参见（Garfield，1979）。

利用档案资料研究网络的例子绝不仅仅局限于上面两个例子。其他例子还有很多，如：

（1）校际交流、E-mail联系、微信中的关系等；

（2）国际贸易关系、制造业交易、原材料往来；

（3）国际政治关系、敌视关系等；

（4）移民记录；

（5）合资企业、高校之间的合并；

（6）连锁董事会。

（五）观察法、实验法

除了上述方法之外，我们还可以利用观察法（Mitchell，1969）、实验法收集整体网络资料，尽管这些方法使用得较少。观察法的优点是可以观察到自然发生的情形，可以在不干扰被调查者的情况下获得准确的数据；缺点是耗费时间长，仅适用于观察小规模的社会互动场景。

观察法又可以分为两种：一种是在实验室内安排一个观察者，让他记录下来眼前发生的所有互动。另外一类是时间分配方法（time allocation method），这种方法主要应用于“个体生态学”（ethology），观察者随机地出现在各种场合之中，记录下来在一个很短时间内哪个人与哪个人之间存在交往，交往的内容是什么等。

实验法也可用来收集整体网资料。人工实验法仅适用于小规模的社会互动研究，而计算机模拟实验则可以分析大规模的整体网数据。

总的来说，实验法在有关二方组、三方组的研究中利用得较多。在过去的几

十年时间里,很多学者研究了学校之间信息的传播问题。利用人工实验法的一个著名研究是小世界研究(Travers & Milgram, 1969),其中斯坦利·米尔格兰姆(Stanley Milgram)在一系列研究中探讨了在美国任意选择出来的两个人之间需要经过多少步骤就可以建立联系,该研究发现了六度分割(six degrees of separation)命题,认为全世界的人只需要通过六步就可以建立联系。该发现对当代主要利用计算机模拟实验法进行的复杂网络研究有重要的启发(Watts, 2003)。该项研究也成为《六度分割》(*Six Degrees of Separation*)这个剧本和网络凯文·培根游戏(Kevin Bacon Game)的来源(后来成为由威尔·史密斯主演的电影)。该游戏可以找到一个演员(比如成龙、章子怡等)通过几个中间人建立与凯文的联系。

另外,也可以用实验法检验网络结构对行动者的社会行为的影响。例如,贝弗拉斯(A.Bavelas)于 20 世纪 50 年代进行了一次实验,分别检验了由五个人在有关如何解决问题的信息交往方面构成的"星形结构"、"链形结构"、"环形结构"和"Y 形结构"对问题的成功解决产生什么影响。这些结构如图 2.2 所示,四种图形分别代表所研究的四种交往模式。贝弗拉斯小组针对主要自变量——中心度,提出一种正式的模型。在研究的时候,不让参与者知道自己所在的实验组具有怎样的结构(参见 Everett, 2002:56—59; Freeman, 2004:70—71)。

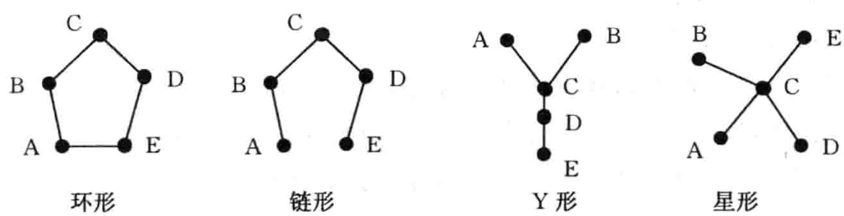


图 2.2 贝弗拉斯小组研究的四类网络结构

研究结果显示,这四种结构对行动效果,即解题效率的影响各不相同。"环形结构"和"Y 形结构"比"星形结构"和"链形结构"在解题时间方面更快。这种研究展示了"结构"对"行动"的影响,是一种社会心理学的实验研究,与我们所说的严格意义上的整体网研究不同。

不管怎样,贝弗拉斯的研究激发了广泛的兴趣。来自许多领域的学者相继利用实验法开展研究,扩展贝弗拉斯的工作。但是,这种兴趣迅速消退,因为贝弗拉斯于 1950 年从政。"随后的探索者或者试图关注中心度的概念问题,或者

关注交往结构对问题解决的影响,而不是二者兼而有之。”(Freeman, 2004:71)每个新研究者都关注整个社会网络图景的局部,失去了其全景性。尽管如此,直到当代实验法仍然被网络学者使用。

(六)问卷法

利用问卷法收集整体网资料的方法有多种。第一种是所谓的“以行为为基础”的方法,即针对整体的每一位行动者进行问卷调查,使得每份问卷都在整体网的邻接矩阵中形成一行。例如,在收集七个人之间的建议关系的时候,可形成下面的邻接矩阵。交给 Andy 的建议关系问卷产生了第 1 行数据,交给 Bill 的问卷产生第 2 行数据,依此类推。可见,Andy 在自己的问卷中选择了向 Bill、Dan 和 Garth 提供建议。这意味着,尽管我们把一个整体网看成是一个整体,但是其中的每行都有不同的来源,可能拥有自己的偏差。参见表 2.4。

表 2.4 七人关系矩阵

	And	Bil	Car	Dan	Ele	Fra	Gar
Andy		1	0	1	0	0	1
Bill	1		1	0	1	0	0
Carol	1	1		1	1	0	0
Dan	1	1	1		0	0	0
Elena	0	0	0	0		1	0
Frank	0	0	0	0	1		0
Garth	1	1	0	0	0	0	

第二种是“以行和列为基础”的方法,要求每个人既说出自己给谁提供建议,也指出自己得到谁的建议。也就是说,对于任何两个人如 Andy 和 Bill 来说,我们拥有两个数据点:来自 Andy 的一个数据点和来自 Bill 的一个数据点。然后,我们需要某种用来判断最终数据矩阵中数字的规则。例如,如果二者之一说向对方提供了建议,那么矩阵相应的值就是 1 吗?还是二者都说自己得到了对方的建议,才用 1 来表示?这个标准需要研究者自己来决定。

第三种方法是斯蒂芬·薄加提(Steve Borggati)和大卫·克拉克哈特(David Krackhardt)提出来的“一致性方法”(consensus method),这需要询问群体中的每个成员,让他们说出群体中任何两个人之间的关系。其结果是,对于每一对人,如 Andy 和 Bill 来说,我们有 N 个数据点,此时我们也需要某种标准,用来确定“正确”的答案。参见本书第十章。

（七）导出法

假设已知行动者—事件或者行动者—群体数据，我们就能够据此构造一个行动者—行动者(actor-by-actor)矩阵，这需要计算每一对行动者共享的事件/群体的数量。例如，在 UCINET 自带的数据库文件夹中，有一个数据矩阵名为 davis，它是戴维斯等学者(Davis, Gardner & Gardner, 1941)于 20 世纪 30 年代通过分析报纸上的社交信息，记录下来 18 位女性参加哪些社交事件的情况所形成的矩阵，数据如表 2.5 所示。

表 2.5　18 位女性参与社交矩阵

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Evelyn	1	1	1	1	1	1	0	1	1	0	0	0	0	0
Laura	1	1	1	0	1	1	1	1	0	0	0	0	0	0
Theresa	0	1	1	1	1	1	1	1	1	0	0	0	0	0
Brenda	1	0	1	1	1	1	1	1	0	0	0	0	0	0
Charlotte	0	0	1	1	1	0	1	0	0	0	0	0	0	0
Frances	0	0	1	0	1	1	0	1	0	0	0	0	0	0
Eleanor	0	0	0	0	1	1	1	1	0	0	0	0	0	0
Pearl	0	0	0	0	0	1	0	1	1	0	0	0	0	0
Ruth	0	0	0	0	1	0	1	1	1	0	0	0	0	0
Verne	0	0	0	0	0	0	1	1	1	0	0	1	0	0
Myrna	0	0	0	0	0	0	0	1	1	1	0	1	0	0
Katherine	0	0	0	0	0	0	0	1	1	1	0	1	1	1
Sylvia	0	0	0	0	0	0	1	1	1	1	0	1	1	1
Nora	0	0	0	0	0	1	1	0	1	1	1	1	1	1
Helen	0	0	0	0	0	0	1	1	0	1	1	1	1	1
Dorothy	0	0	0	0	0	0	0	1	1	1	0	1	0	0
Olivia	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Flora	0	0	0	0	0	0	0	0	1	0	1	0	0	0

在矩阵中，行代表每一位女性的姓名，列代表她们参与的各个事件。我们可以据此构建一个女性—女性矩阵，构造的方法有两种。一种方法是利用 UCINET 中的矩阵算法(Matrix Algebra)，计算该矩阵与其转置矩阵之积($Y=XX'$)。计算的结果是一个矩阵，其第 ij 项值记录的恰恰是女性 i 和女性 j 共同参加的事件数目。要想实现这个想法，需要在 UCINET 中执行的计算步骤有：首先对 davis 矩阵进行转置处理(沿着 Transform→Transpose 路径，在相应的

文件夹中选出 davis 矩阵即可得到该矩阵的转置矩阵，可命名为 davisT)，然后计算 davis 矩阵与其转置矩阵 davisT 之积，令其等于 Y，即 Y＝davis * davisT（沿着 Tools→Matrix Algebra 路径，点击后会出现“Enter command”（键入命令）对话框，在该对话框中键入 Y＝prod(davis, davisT)，如图 2.3 所示。

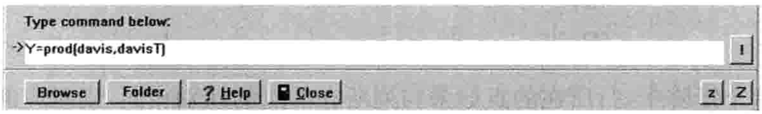


图 2.3 矩阵计算命令

回车之后，就可以计算出“女性—女性矩阵”，如表 2.6 所示。

表 2.6 女性共同参与社交矩阵

	Eve	Lau	The	Bre	Cha	Fra	Ele	Pea	Rut	Ver	Myr	Kat	Syl	Nor	Hel	Dor	Ola	Flo
Evelyn	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
Laura	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
Theresa	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
Brenda	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
Charlotte	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
Frances	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
Eleanor	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
Pearl	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
Ruth	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
Verne	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	3	1	1
Myrna	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	4	1	1
Katherine	2	1	2	1	0	1	1	2	2	3	4	6	6	5	5	4	1	1
Sylvia	2	2	3	2	1	1	2	2	3	4	4	6	7	6	6	4	1	1
Nora	2	2	3	2	1	1	2	2	2	3	3	5	6	8	6	3	2	2
Helen	1	2	2	2	1	1	2	1	2	3	3	5	6	6	7	3	1	1
Dorothy	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	4	1	1
Olivia	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
Flora	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

利用同样的方法，可以把一系列坐标点转变为各个“点对”之间的距离矩阵，可以把诸如性别这样的二值数据转变为“二人组性质的数据”(dyadic attribute data)，如“与之属于同一性别”。例如，下面五个人的性别之间的关系即可列表为：

表 2.7 五人之间"性别关系"矩阵

	M	F	F	M	F
Male	1	0	0	1	0
Female	0	1	1	0	1
Female	0	1	1	0	1
Male	1	0	0	1	0
Female	0	1	1	0	1

在这个矩阵中,行所在的点如果与列所在的点的性别相同,则对应的矩阵值为 1,否则为 0。

第二种方法比较简单。要从 2-模网络中构造 1-模网络,只要在 UCINET 中利用 Data→Affiliations(2-mode to 1-mode)程序,选择要分析的 2-模数据计算即可。该方法可以计算出两个 1-模矩阵,分别对应于行和列的 1-模矩阵,这里不再展示具体步骤。

关于整体网研究,特别是整体网资料的收集方法,请读者进一步参见罗家德(2005:63—89)。

二、问卷设计应注意的问题

整体网的问卷设计也遵循在一般问卷设计中所坚持的原则,但稍有不同。整体网问卷的设计已经在有关文献中得到了详细的阐述(罗家德,2005),请读者参考该书第四章。这里择其要述之,并力图有所补充。

整体网研究往往是先具有较明确的边界,一般不能通过随机抽样的方法得到样本,分析单位的选取也比较明确。整体网研究一般利用方便抽样方法得到样本,不使用随机抽样方法得到样本。运用随机抽样方法得到的样本往往只能进行个体网研究,不能进行整体网研究,因为进行整体网研究没有什么实际意义。

另外,由于在一对行动者之间可能存在多类关系,因此,整体网数据的收集往往从多维关系的角度进行,即同时了解行动者之间的多种关系。这里涉及一些调查技巧,如问卷尽量不署名,但可以事后由研究者自己加上被调查者,即问卷填写者的姓名;保证不泄密;强调本研究纯属学术研究。得到资料之后,需要输入计算机,其过程参见本书的数据输入部分以及罗家德(2005:71—77)。

在调查整体网中各个行动者之间的某种具体关系的时候,要求每个被调查

者说出自己与所有其他行动者之间的关系是否存在，如果存在，其紧密程度如何。例如，如表 2.8 所示，在考察五个“国家”之间的“出口贸易关系”的时候，需要由研究者根据既定的数据判断两个国家之间的关系是否存在，如果存在关系，强度如何。如果国家 1 向国家 2 出口，则令第一行第二列的元素值为 1。当然，这种数据也可以是多值关系数据，可以把出口贸易额放在各个相应元素值上。

表 2.8　数据矩阵的模板

行动者姓名	1	2	3	4	5
1					
2					
3					
4					
5					

三、整体网资料的整理和分析

用来分析、整理整体网数据的方法和软件很多。例如，在“国际社会网络分析网”(INSNA)网站(www.insna.org)上可以链接到的网络软件达到几十种。其中，UCINET 是最常用的一种。另外，很多研究需要特定的程序进行数据分析，因此，很多学者要自己编写程序来分析网络数据(如网络变量的统计推断研究，p 模型研究等)，这种研究的要求较高。一般情况下，我们可用 Excel 软件首先输入矩阵数据，然后用 UCINET 软件进行转换并分析。数据转换的具体方法不再赘述。

数据输入后，需要进行统计分析。用来分析整体网数据的统计方法不是通常意义上的线性相关分析、偏相关分析、方差分析、回归分析等，这些常规的统计方法一般不能用来分析整体网络资料。那么，整体网研究运用哪些统计分析方法呢？回答是：这取决于研究者分析整体网的哪个层次，不同层次的整体网研究对应着不同的分析方法。如果研究“二方关系”层次的整体网络结构，可以根据特定的计算程序，计算出二方谱系，对有关二方关系的命题进行检验。如果研究“子群”层次的整体网络结构，可以利用 UCINET 软件，计算出“派系”、“n-派系”、“k-丛”等。如果研究动态网络，往往需要利用计算机模拟方法(常见的模拟软件有 Repast，Swarm 等。汤姆 · A.B.斯奈德(Tom A.B.Snijders)教授等人编

写的 Siena 程序则专门分析网络数据的演变);三方关系研究则往往需要研究者自己编写独特的相关程序。这些内容比较复杂,有关具体的分析步骤从略。

四、整体网资料的相关问题

(一) 信息的精确性

在调查的时候,被调查者能否真正告知自己的社会网络情况?这是需要深究的一个问题。很多营销学者发现,消费者很少会告诉你昨天午餐吃的是什么。贝尔纳等学者(Bernard, Killworth & Sailer, 1982)系统地探讨了信息的精确度问题,发现被调查者所说的信息 52%是错误的。研究发现,人们在回忆自己与他人之间的互动时有一种系统的偏向,即偏向于常态而非反常的情形。当然,这种偏差至少要比随机偏差好一些。另外,人们倾向于记住与重要的他者(significant others)之间的互动,倾向于忘记与不重要的他人之间的互动。某些回答者会撒谎,目的是为了让研究者认为自己是较好的。

在问卷中也存在如何解释概念的问题。"朋友"对于不同的人来说含义是不同的。克拉克哈特(D. Krackhardt)的解决办法是,让被调查者自行判断自己与他人之间是否为朋友关系。因此,如果某人声称自己与每个人都是友谊关系,而他人却都声称自己没有朋友的话,那么我们就得到一个启示,即有人可能在撒谎,或者误解了问题。

(二) 抽样中的问题

前文指出,我们一般不能利用抽样方法来研究整体网。然而,可以利用抽样资料对某些假设进行检验。例如,可以根据抽样得到的点之间的关系资料来估计一个网络的密度,但这里是有问题的。如果从总体为 10 000 人的社区中随机抽出 200 人进行整体网调查,那么这 200 人之间的关系密度在多大程度上代表整体网络的密度?这是难以计算的。个体网资料可以进行推断,整体网资料则很难进行推断。

第三节　整体网研究的优势及局限

任何研究范式都不是万能的,都有其适用范围和优缺点,整体网研究也不例外。如果说个体网研究得到的结论具有推断意义的话(因为个体网数据往往是根据随机抽样方式得到的),那么就整体网的研究结论来说,由于多数整体网的

资料都是通过"方便抽样"(convenient sampling)得到的，因此，整体网研究得到的结论往往不具有统计推断的意义，其结论只适用于所研究的群体(罗家德，2005:40—41)。这是必须强调的。

但是，整体网研究的结论往往具有很大的参考价值。这里有两点需要强调。首先，有的研究的目的本身就是为了描述现象、揭示整体网的结构，而不是为了推断。例如，研究一个组织内部人际关系网络的目的可能仅仅是为了找到"地下司令"是谁，找到哪些人构成小派系，而不是为了找到具有推广价值的结论。在这种情况下，整体网研究就派上了用场，学者也不关注"统计推断"问题。其次，如果研究的目的是为了推广的话，那么研究的结论到底在多大程度上可以推广？我们往往不能给出量化的指标来，因为整体网数据不是利用随机抽样方法得到的。尽管如此，由于一定地区的行动者之间的关系模式往往具有不同程度的"共性"，遵循一定的"模式"(pattern)，因此，我们可以说也只能说，整体网研究的结论具有"一定的代表性"，而不能指明在多大程度上具有代表性。例如，就"亲属关系"结构来说，"东北地区"的亲属称谓、亲属结构具有很大的相似性。考虑到这一点，可以认为由一个村落的亲属关系结构可以"在一定程度上"推论到其他村落的结构，尽管这种推论不是统计意义上的。

一、整体网研究的优势

整体网研究有助于揭示整个网络的各种结构特征。例如，可以计算出网络的互惠指数、揭示网络中的三方关系结构以及整体网络结构，找到整体网络中的分帮分派情况，计算出网络的密度等。显然，这种研究是个体网研究所不能达到的。针对一个群体进行整体网研究，找到其中的非正式小群体，可以为高校、企业、政府、或者非政府组织等搞好人际关系、提高效率提供参考。

整体网研究的另外一个优势在于，它允许同时把社会系统视为一个整体和构成更大整体的某个部分。这种研究可以揭示整体系统的整合性，发现整体网的层次性、等级性和阶层性等，可以解释联系的紧密性与整体网成员的行为之间的关系，可以找到整体网中联络和分解的模式，找到结构对等的行动者等。

二、整体网研究的局限

1. 关注整体，看不到个体网的各种特征。这恰恰是个体网研究的优点所在。当然，在对整体网的结构进行描述和解释时，可以结合个体行动者的属性数

据,就像可以根据个体行动者的特点来解释块模型那样。尽管如此,抽象的整体网结构研究确实有仅仅关注“形式”,忽视“内容”的危险。

2. 整体网结构隐含着怎样的行为规范?人们为什么要互惠?这些问题是进行量化的整体网研究所不能回答的。正是在这个意义上,我们认为个体网研究、整体网研究应该结合起来;规范研究和形式研究结合起来,这样才能更好地描述、解释社会行为。

3. 整体网研究在方法论上不总是可行的,在分析上也不总是令人满意。这一方面是因为进行整体网研究必须首先规定整体的边界,列举出整体中的全部成员名单,调查他们之间的各种关系。这些显然是存在问题的。例如,边界是明确的吗,不正确的边界会导致分析的混乱。另外,整体网全部成员之间的所有关系是很难调查到的。

总之,整体网研究的目的往往不是为了推广,毋宁说是为了揭示其结构。鉴于不同类型的结构对行为有重要影响,因此这种整体网络结构研究具有重要意义。

最后需要补充的是,在当代社会网络研究领域中,取得重要突破性进展的领域和方向多数是在整体网研究中出现的。例如,布雷格(Breiger, 2003)指出,40年以来,社会网络研究集中于如下几个方面:

(1) 对点、线、整体网的研究,例如研究中心度、中心趋势等,对各种限制性因素进行测量(Burt, 1992)。这既是关于个体网方面的研究,也是整体网研究的内容。

(2) 研究多元关系网络,对角色关联进行模型分析。有两位学者(Lazega & Pattison)对一家美国公司的三个办公室之间的工作关系、提供建议关系和朋友关系进行了富有创造性的统计模型研究。

(3) 对“对等性”(equivalence)的研究也是重点之一。此类研究把个人层次数据和整个网络的宏观结构连在一起。此类研究主要有三类:“结构对等性”、“自同构对等性”和“规则对等性”。除此之外,学者们又提出了具有一般形式的对等性研究(Borgatti & Everett, 1992a, 1992b, 1992c)。对网络的宏观结构模型研究更得到长足的进展(Nowicki & Snijders,2001)。

弗里曼(Freeman, 2004)进一步指出,当代社会网络研究有如下四个特点:

(1) 对社会行动者之间的某种特定关系的结构研究;

(2) 建立在系统的数据基础上;

(3) 大大依赖于图论语言和技术；

(4) 应用数学模型、统计技术和计算机模拟技术。

整体网研究的进展要比个体网研究快得多，并且在国外已经取得很多重要成果。有关研究已经非常深入，模型很复杂。实际上，整体网研究的学者往往专攻于一种十分具体的网络模型（如块模型、核心—边缘模型等），并在该网络模型领域不断推进下去。从这个意义上说，我们很难对各个网络模型都有非常透彻的理解和掌握。因此，该领域的前沿性工作需要不断地学习和研究。

本 章 小 结

本章首先探讨了整体网研究的一些基本内容，包括整体网的分类、规模、构成、成员之间的距离等，特别介绍了各种整体网的结构研究。然后介绍整体网资料的收集方法，包括线人法、提名法、职位生成法、档案、观察和实验法等。整体网资料的整理和分析需要利用专门的软件，其研究结果往往不具有统计推断价值，但是这并不妨碍整体网研究的意义所在，即这种研究可以揭示整个网络的结构特征，可以为现实生活中的各种关系研究提供学理支持和实证参考，这是其优势所在。整体网研究的发展离不开电脑编程、统计分析技术的突破，因而，这种研究越来越复杂。整体网研究的局限性在于，它不关注网络内含的文化规范和个体网的性质。因此，整体网研究的结果要结合个体数据和文化要素进行解释。

第三章

整体网分析工具——UCINET 简介*

可用来分析整体网的软件很多，本章介绍其中的一种——UCINET 软件。首先介绍其运行环境，然后介绍数据输入和分析的方法。

第一节 UCINET 的运行环境

UCINET(University of California at Irvine NETwork)是一种功能强大的社会网络分析软件，它最初由社会网络研究的开创者、加州大学欧文分校(University of California at Irvine)的林顿·弗里曼(Linton Freeman)教授编写，后来主要由新一代学者、目前分别供职于美国肯塔基大学(University of Kentucky)的斯蒂芬·博加提(Steve Borgatti)和英国曼彻斯特大学(University of Manchester)社会科学学院的马丁·埃弗里特(Martin Everett)维护更新。该软件可从网站(http://www.analytictech.com/downloaduc6.htm)上免费下载使用一个月，在此感谢作者的慷慨。UCINET 软件只占用不到 100 M 的空间，下载后可直接点击安装，具体安装过程很简单，这里从略。安装后，UCINET 中包含一个用户手册(User's Guide)。

UCINET 包括大量的网络分析指标，如中心度、二方关系凝聚力测度、位置分析算法、派系分析、随机二方关系模型(stochastic dyad models)p^1 以及对网络假设进行检验的程序(包括 QAP 矩阵相关和回归、定类数据和连续数据的自相关检验等)等，还包括常见的多元统计分析工具，如多维量表(multidimensional scaling)、对应分析(correspondence analysis)、因子分析(factor analysis)、聚类分析(cluster analysis)、针对矩阵数据的多元回归(multiple regression)等。除此之外，UCINET 还提供数据管理和转换的工具，可以从图论程序转换为矩阵

---

* 本章是在 UCINET 软件中附带的使用手册(2002 版)基础上，根据最新版的 UCINET 6.504(2014 年 1 月 16 日)，结合笔者的理解撰写而成。

代数语言。

在 UCINET 6 中,全部数据都用矩阵形式来存储、展示和描述,请读者注意这种矩阵数据的形式。UCINET 软件可以处理 32 767 个点的网络数据。下面介绍 UCINET 的菜单、格式、UCINET 6 数据库的实质、执行分析和输出结果等。由于该软件在不断更新,因此其菜单也时常有变化,这里结合 2014 年 1 月 16 日的最新版本 UCINET 6.504(即修改了 504 次的 UCINET 6 版本)来介绍其应用。在安装完毕之后,打开软件出现的主界面如图 3.1。

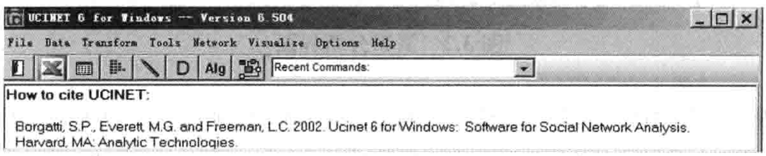


图 3.1 UCINET 主界面

UCINET 主界面中包括主菜单和子菜单,子菜单还可能包含子—子菜单。在主菜单最顶端包含八个选项(File, Data, Transform, Tools, Network, Visualize, Options, Help),下面将分别介绍。八个主菜单下面有八个快捷图标,点击后将直接进入相应的程序,免去了中间环节。从左到右,这八个快捷图标分别表示(1)Exit:退出 UCINET 程序;(2)Excel Matrix Editor:Excel 数据矩阵编辑器;(3)Matrix Spreadsheet Editor:建立 UCINET 本身的矩阵数据表;(4)DL-Editor--Import text data from spreadsheet,数据语言编辑器,即导入文本文件,或者从 Excel 中导入文件,并保存为 UCINET 格式的数据;(5)Edit Text File:编辑文本文件;(6)Display UCINET Dataset:展示 UCINET 数据;(7)Alg:直接调用矩阵算法(Matrix Algebra);(8)Visualize network with NetDraw,导入 NetDraw 可视化程序,对网络数据进行可视化分析。下面首先分析 UCINET 的数据格式,然后介绍菜单的功能。

一、UCINET 的数据分析格式

(一) 参数格式

假设你从菜单中选择了一项程序,除了少许的例外,下一步工作通常是考察"参数格式"(parameter form)。一个格式是由一个或多个填空式问题(fill-in-the-blank questions)构成的。例如,假设在 UCINET 主菜单中依次点击了 Net-

work→Subgroups→Cliques,你看到的格式就由图 3.2 左侧这八个问题构成。

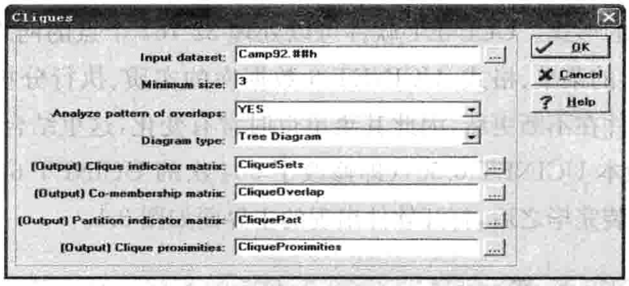


图 3.2　派系分析界面

左侧的八行内容(黑体字)是程序提出的问题,右侧包含一些相应的回答,其中"Input dataset"要求输入数据,要求研究者在相应的硬盘中找出需要加以分析的数据,例如右侧的 Camp92.##h 数据文件。其余七个都是系统默认的回答(default answers)。如果这些满足你的要求,就可以点击 OK 以便执行分析。

(二) 执行分析

UCINET 6 在执行分析的时候,要接收数据文件,并输入一定的参数。分析结果输出的文件有两种,分别为文本文件和数据文件。

首先看 UCINET 6 的数据文件。多数程序只接受单个数据作为输入文件。当然,该文件可能包含多个矩阵,可以表达同一组人之间的多种关系。少数程序则把多个数据库作为输入文件。例如,Data→Join 命令的含义是将两个或多个数据文件合并成为一个数据文件,可针对行或列合并,也可以合并成由两个或多个矩阵组合在一起的多元矩阵。Tools→Testing Hypothesis→Diadic(QAP)命令则需要两个或多个匹配的数据矩阵,计算它们之间的相关或回归关系。

其次,在计算的时候往往需要研究者输入一定的参数。例如,派系分析(Network→Subgroups→Cliques)程序需要输入的参数包括:数据文件名、派系成员的最小数目、各种输出的文件名等。

输出文件名通常是默认值,可针对该输出文件作进一步统计分析。例如,派系分析(Cliques)程序的一个输出文件是行动者—行动者矩阵(actor-by-actor matrix),其第 ij 项元素给出的是行动者 i 和行动者 j 共同参与的派系的数目。该矩阵还可用作输入矩阵,进而进行多维量表分析(MDS program),以便获得对叠加模式的空间表达。因此,该输出数据(Co-membership mtrix)的默认名为

CliqueOverlap。

除了输出的数据文件之外，大多数 UCINET 程序还生成一个文本文件作为输出文件。该文件名总是存为日志文件(Log File)，并且在屏幕上自动显现出来。

还是以上述派系分析程序为例。第一步是选择该程序，即选择 Network→Subgroups→Cliques，会出现对话框。第一行要求指定包含数据的 UCINET 6 数据库名。如果数据文件名不是 Camp. ##H，而是位于 C 盘的\UCINET 6\DataFiles 文件夹中的 TARO，则需要键入(或者点击其右侧"三点按钮"处选出)的文件名为 C:\UCINET 6\DataFiles\TARO. ##H。

第二行要求输入派系的最小规模，默认值是 3。如果是分析规模为 4 的派系，则需要将"Minimum size:"后面的 3 修改为 4。该数字越小，分析得到的派系数量也就越少。

第三行询问是否计算关系重叠的模式，即分析一个行动者—行动者矩阵(actor-by-actor matrix)，从而计算出每一对行动者之间的共同隶属的派系的数目，然后把该矩阵提交给等级聚类分析(hierarchical clustering)。默认回答是 YES。当然，在处理特别大的数据的时候，为了节省时间，也可以选择 NO，它只分析派系，计算每个行动者的派系参与值。

第四行追问的问题是，你需要看什么样的聚类图(clustering diagram)，默认值是树形图(tree diagram)。另外一类图是树状图(dendrogram)。无论是树形图，还是树状图，二者在描述网络关系的结构方面是完全一致的。

第五行到第八行给出的是默认的输出文件名。这些文件可用作进一步分析。输入这些参数之后，点击 OK 后得到的分析结果如表 3.1 所示(节选)：

表 3.1　派系分析结果

```
10 cliques found.

  1:  2 3 17
  2:  1 2 17
  3:  17 18 22
  4:  4 5 6
  5:  4 6 7
  6:  5 20 21
  7:  8 9 10
  8:  11 20 21
  9:  12 13 14
 10:  12 14 15
```

由此结果可见，数据中包含的规模为 3 的派系共有 10 个。例如，第 2 派系的成员为 1, 2, 17，第 3 派系的成员为 17, 18 和 22。派系的成员之间可以重叠。

下一步给出了一个行动者—行动者派系共享成员矩阵(an actor-by-actor

clique co-membership matrix),该矩阵给出了每一对行动者共同隶属的派系的数目,如表 3.2 所示。

表 3.2 行动者—行动者共享派系矩阵

Actor-by-Actor Clique Co-Membership Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
3	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4	0	0	0	2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
6	0	0	0	2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0
12	0	0	0	0	0	0	0	0	0	0	0	2	1	2	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	2	1	2	1	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0
21	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1

由这个结果可见,第 17 行第 18 列的值是 1,这说明 17 号行动者和 18 号行动者隶属于一个派系,由第一个分析结果可以看出他们隶属第 3 个派系。而第 2 行第 17 列的值是 2,这说明 2 号行动者和 17 号行动者隶属于两个派系,由第一个分析结果可以看出他们隶属于第 1 和第 2 个派系。对角线上的值表示的是行动者隶属的派系数。例如,第 17 行第 17 列的值是 3,这说明该行动者参与了三个派系的活动。

这个矩阵是程序自动给出的聚类分析表的基础,该表对"共享成员"矩阵进行了"等级聚类"(hierarchical clustering)分析,结果如表 3.3 所示。

表 3.3 重叠矩阵的等级聚类分析

```
HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

            1 1 1 1 1 1 1         1 2 2       1 1 2
Level   8 9 0 3 2 4 5 6 9 5 6 4 7 1 1 0 3 1 2 7 8 2
-----   - - - - - - - - - - - - - - - - - - - - - -
2.000   . . . . XXX . . . . XXX . . XXX . . XXX . .
1.000   XXXXX XXXXX . . . XXXXX . XXXXX . XXXXX XXX
0.750   XXXXX XXXXXXX . . XXXXXXX XXXXX XXXXXXX XXX
0.400   XXXXX XXXXXXX . . XXXXXXX XXXXX XXXXXXXXXXX
0.080   XXXXX XXXXXXX . . XXXXXXXXXXXXX XXXXXXXXXXX
0.000   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

该结果给出了 22 个行动者共同隶属于派系的聚类分析表。由此可以看出,12 号和 14 号、4 号和 6 号、20 号和 21 号、2 号和 17 号都共同参与了两个派系的活动。

还有一个问题是，每个派系包含多少个人？两个派系“共享”的人有多少？因此，给出的下一个计算结果就是派系—派系共享成员矩阵(clique-by-clique actor co-membership matrix)，该矩阵给出了每一对“派系”共同拥有的行动者数目，如表 3.4 所示。

表 3.4　派系—派系共享成员矩阵

Clique-by-Clique Actor Co-membership matrix

	1	2	3	4	5	6	7	8	9	10
1	3	2	1	0	0	0	0	0	0	0
2	2	3	1	0	0	0	0	0	0	0
3	1	1	3	0	0	0	0	0	0	0
4	0	0	0	3	2	1	0	0	0	0
5	0	0	0	2	3	0	0	0	0	0
6	0	0	0	1	0	3	0	2	0	0
7	0	0	0	0	0	0	3	0	0	0
8	0	0	0	0	0	2	0	3	0	0
9	0	0	0	0	0	0	0	0	3	2
10	0	0	0	0	0	0	0	0	2	3

该结果给出了 10 个派系共同拥有的行动者的情况。对角线表示每个派系都拥有三个成员(这是系统默认的派系规模)。第 1 行第 3 列的值 1 表明，这两个派系共享的行动者有 1 个，由第一个分析结果可以看出共享的行动者是第 17 号行动者。

针对“派系共享成员”矩阵进行聚类分析的结果如表 3.5。

表 3.5　派系共享成员矩阵的等级聚类分析

```
HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

                             1
Level   1 2 3 7 4 5 6 8 9 0
-----   - - - - - - - - - -
2.000   XXX . . XXX XXX XXX
1.000   XXXXX . XXX XXX XXX
0.250   XXXXX . XXXXXXX XXX
0.000   XXXXXXXXXXXXXXXXXXX
```

该结果给出了 10 个派系共享的行动者的聚类分析表。可以看出，派系 1 和派系 2 共享 2 个行动者。

该程序还生成一个聚类图(clustering diagram)，如图 3.3 所示。

从图中可以看出，12 和 14 共同隶属于 2 个派系，4 和 6，21 和 20，2 和 17 也都共同隶属于 2 个派系。而 8，9，10 隶属于 1 个派系。13，12，14 隶属于 1 个派系。5，6，4 隶属于 1 个派系。可以看出，16 和 19 都不隶属于任何派系。可见，该图清楚地告诉了我们派系分派的情况。

此外，输出结果还表明了程序生成了哪些文件。例如，有一个输出文件是派系—行动者二元矩阵(group indicator matrix)，它表明哪个行动者隶属于哪个派系，并被程序命名为 CliqueSets。需要注意的是，该结果并不展示在计算结果

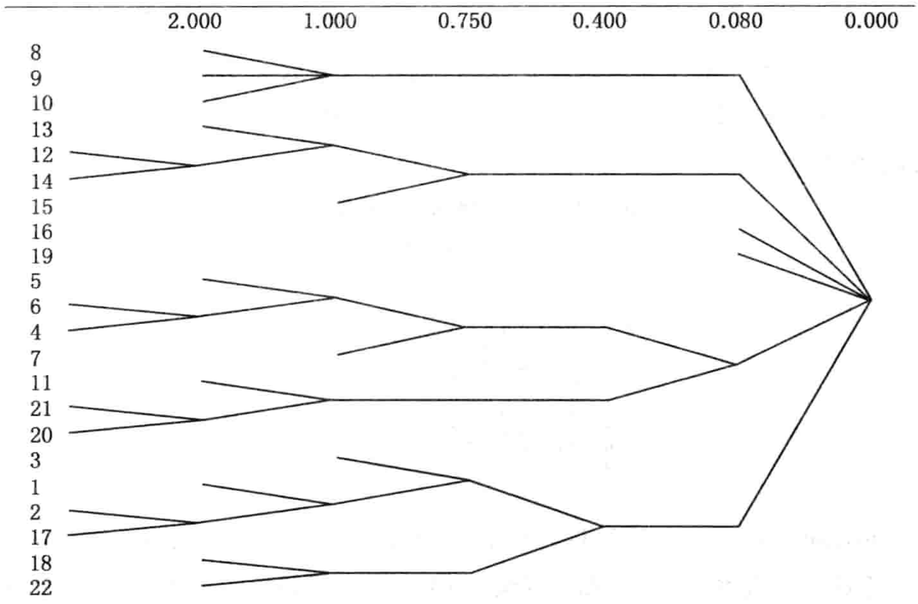


图 3.3　聚类图

当中,而是自动保存在默认的文件夹中。如果分析人员想看这个矩阵,可以点击 Data→Display,在数据所在的文件夹中选出 CliqueSets 打开即可,如表 3.6 所示。

表 3.6　隶属关系矩阵

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0
4	0	0	0	1	1	0	0	0	0	0
5	0	0	0	1	0	1	0	0	0	0
6	0	0	0	1	1	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	0	1	1
13	0	0	0	0	0	0	0	0	1	0
14	0	0	0	0	0	0	0	0	1	1
15	0	0	0	0	0	0	0	0	0	1
16	0	0	0	0	0	0	0	0	0	0
17	1	1	1	0	0	0	0	0	0	0
18	0	0	1	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	1	0	1	0	0
21	0	0	0	0	0	1	0	1	0	0
22	0	0	1	0	0	0	0	0	0	0

我们可以对该矩阵作进一步分析，例如可用 Data→Extract 子菜单从该矩阵中截取出一个子图。如果你希望从前四个派系中取出朋友关系网，只需要分别点击 Data→Filter/Extract→Submatrix 程序即可，但是要告诉程序初始数据是 CliquesSets，保留全部行，仅选择 1，2，3，4 列（即在“Which columns：”项键入 1，2，3，4 即可），如图 3.4 所示。这样就可以选出前四个派系的情况。

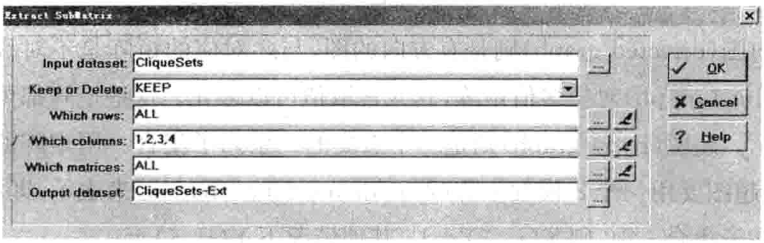


图 3.4　提取子矩阵对话框

上例展示了 UCINET 的输入和输出结果。大多数 UCINET 的分析程序的输入方式和输出文件都与此类似。首先汇报参数设置，然后汇报发现的结果（如派系数）等。

（三）日志文件

在数据分析的过程中，我们可以重新查看刚刚进行分析得到的日志文件（Log Files）及计算结果（也就是说，如果退出了 UCINET，这些日志文件就消失了），查看该文件的方式是点击 File→View Previous Output。在默认状态下 UCINET 可以保存 50 个以前的文件。需要注意的是，一旦退出了 UCINET，再打开 UCINET 时，此前分析得到的日志文件记录都将消失。当然，计算得到的结果还保存在相应的文件夹中。

（四）数据格式——矩阵

关于 UCINET 6 的数据，有三点需要注意。

第一点，关系数据都用矩阵形式保存，也可画图表示。图有多种，如无向图、有向图、多值图等。在具体计算的时候，需要将图的信息转化为矩阵中的数字信息。例如，无向关系图（undirected graph）则可用对称邻接矩阵（adjacency matrix）来表示。表 3.7 的矩阵就是一个邻接矩阵。

表 3.7 五人关系矩阵

	A	B	C	D	E
A	0	1	0	1	1
B	1	0	1	0	0
C	0	1	0	0	1
D	1	0	0	0	0
E	1	0	1	0	0

有向图(directed graph)则是有方向的图,与之对应的矩阵是不对称的。多值图(valued graph)对应多值矩阵,其元素取值可以多元。这些矩阵都可用 1-模方阵表示。超图(hypergraph)则是一个点集的一系列子集合。用矩阵的形式来说,一个超图要用一个 2-模发生阵(2-mode incidence matrix)来表示,即如果点 i 被包含在子集合 j 中,则 $Y(i, j)=1$,其他情况下 $Y(i, j)=0$。

尽管 UCINET 中的数据都是矩阵,但是矩阵未必都表征网络。例如,如下三组数字都表示矩阵,但矩阵 3 就是一个数字罢了,不表示网络:

矩阵 1	矩阵 2	矩阵 3
2 3 4	3	3.14
1 0 8	4	
	6	
	0	

另外,一个 UCINET 数据文件可能包含多个匹配的矩阵。所谓匹配,指的是每个矩阵的行和列都必须对应相同的行动者。换句话说,可用一个文件表示一个集合的人之间的多种关系。把多种关系放在一个数据文件中,可进行块模型分析、对等性分析,也可以进行中心性分析。在 UCINET 的矩阵代数算法(Tools→Command Line/Matrix Algebra)中,可对多元矩阵数据进行分析。在这里,矩阵代数程序会把一个多矩阵“思考”为一个单一的三维(3-way)矩阵,由各行、各列和各个层次(即关系的类型)组成的,从而允许使用者在所有这三个维度上执行分析。

第二点需要注意的是,UCINET 数据不是文本文件,不能利用文字处理程序(word processor)来编辑处理。只有 UCINET 才能打开这种数据文件。当然,UCINET 也提供了转换数据的程序(如可转换为 Excel 文件),可以将其他数

据(如 Excel 数据)转换为 UCINET 数据库(参见 Import 命令),反之也可(参见 Export 命令)。

第三点,一个单独的 UCINET 数据库实际上由两个物理文件(physical files)组成。一个包含扩展名“.##D”,它包含关于实际数据的信息;另外一个包含扩展名“.##H”,它包含实际数据。然而,在键入一个 UCINET 数据的时候只涉及.##H 文件,可以利用文件名,或者文件名加上“.##H”,如“Sampson”或者“Sampson.##H”。

二、UCINET 的主菜单功能简介

UCINET 的主菜单中存在诸多功能,有一些功能的含义比较复杂,有的功能即便在 UCINET 本身的 Help 中也没有介绍。以下结合笔者的理解有选择地讲解其内容,仅供读者参考。

(一) 文件操作菜单(File)

在 UCINET 主界面窗口,第一个主菜单就是文件菜单(File),选择“File”菜单后,会出现下拉式的子菜单(如图 3.5 所示),从中选择相应的命令,就可以完成相应的数据处理和操作。下面简单说明文件操作菜单的主要功能。

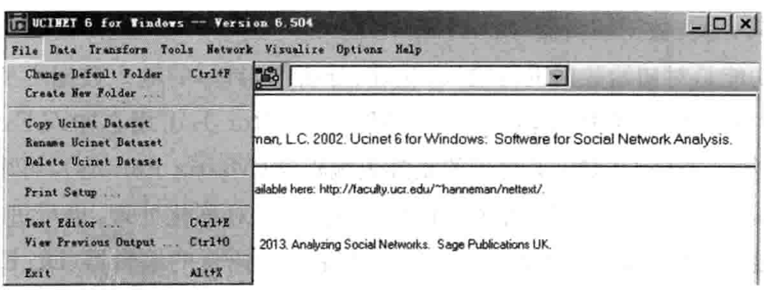


图 3.5　UCINET 的文件命令项

Change Default Folder:改变默认的文件夹。在数据分析的时候,研究者最好将自己经常使用的文件夹设定为默认的文件夹,这样便于快速找到数据,因为 UCINET 是在默认的文件夹中搜索文件的。不管以什么形式,如果你键入有待分析的文件名(如 camp92),UCINET 都假设该文件是在当前路径文件夹中的(current directory),除非你键入全部的路径(full pathname)(如 C:\program files\UCINET 6\datafiles\camp92)。要改变默认的文件夹,需要用到 Change

Default Folder 命令，点击之后，在出现的对话框中依次选出自己经常使用的文件夹即可。

Create New Folder：创建新文件夹。你可以把经常使用的数据放在所创建的新文件夹中，这样的话，只要打开 UCINET，默认的文件夹就是你新创建的文件夹。

Copy UCINET Dataset：拷贝一个已有的 UCINET 数据，再单独起一个文件名。

Rename UCINET Dataset：对已有的一个 UCINET 数据重新命名。

Delete UCINET Dataset：删除已有的 UCINET 数据。

Print Setup：对打印机进行设置。

Text Editor：文本编辑器。可用它来输入、编辑 UCINET 数据文件。

View Previous Output：查看在分析过程中产生的日志文件(log file)。

Exit：退出 UCINET 程序。

（二）数据操作菜单(Data)

数据(Data)菜单包含一些处理 UCINET 数据的命令，它对数据文件进行输入、编辑、输出以及显示分析的结果等，具体包含六类内容。下面简要介绍其含义，其中某些内容的具体介绍可参见本章第三节。

1. 第一类：数据编辑器(Data Editors)。

其功能是对初始数据进行输入、整理、分析等，具体的输入数据类型包括三类：第一类是 Excel 矩阵编辑器(Excel Matrix Editors)，可用它输入 Excel 类型的数据；第二类是 UCINET 自带的矩阵表编辑器(Matrix Editors)，可用它直接输入和编辑 UCINET 数据，加入新的数据表，进行对称化处理，也可进行转置、二值化处理，或者输入随机数据等。第三类是数据语言的编辑器(DL Editors)，它可以用数据语言进行数据的输入，输入的数据类型有十几种。

2. 第二类：数据的导入和导出等命令。

具体又包括如下几类：

(1) Make star graph：创建星形图。所谓星形图指的是一点居于核心，其余点与核心点相连，并且其余点之间无关联的图。点击后，在出现的对话框中可指定星形图的规模、线的方向等。对于无向星形图来说，其对应的矩阵的第一行和第一列都是 1(第一个值除外)，其余值都是 0。

(2) Random：创建随机数据。有时候为了研究的需要，应创建一个满足一

定条件的随机数据，如点出度都是 3 的社群矩阵。可创建的满足一定条件的数据矩阵多种多样，例如创建一个随机的，其元素满足一定概率分布的 UCINET 数据矩阵。可以选择的分布包括一致分布（Uniform）、正态分布（Normal）、二项分布（Binomial）、多项分布（Multinomial）、泊松分布（Poisson）、伽马分布（Gamma）、指数分布（Exponential）等。也可以创建社群矩阵数据（Sociometric）、贝努里分布（Bernoulli distribution）、多项分布（Multinomial）条件的数据，或者满足一定密度条件、元素为 0-1 的 Erdos-Renyi 随机图数据（Erdos-Renyi random graph）。所谓 Erdos-Renyi 随机图，指的是这样的随机生成的图：设定随机图中每一对点之间的边出现的概率相等，并且各条边要独立（Erdös & Rényi, 1959）。吉尔伯特（E.Gilbert, 1959）也独立地提出了这个随机图思想。

（3）Import Excel：导入 Excel 类型或数据语言类型（Data Language, DL）的数据文件，即将它转换为 UCINET 数据。

（4）Import Text File：导入文本文件。可用来导入的文本文件类型包括：数据语言形式的文本文件、多元数据文件（Multiple Data Files）、VNA 软件使用的文件（VNA）、Pajek 软件使用的文件（Pajek）、Krackplot 软件形式的文件（Krackplot）、Negopy 软件使用的文件、ASCII 型的初始文件（Raw）。其中，第一项和最后一项比较常用。

（5）Export：导出文件。把 UCINET 数据导出成为其他类型的文件，以便其他程序读取和分析。可输出的文件类型与输入的文件类型基本相同，但又有所差异，主要包括数据语言文本文件（DL）、Krackplot 文件、Mage 文件、Pajek 文件、Metis 文件、Raw 文件和 Excel 文件。

（6）CSS：输入认知社会结构（Cognitive Social Structure, CSS）类型的文件，其功能是把特定类型的数据转换为标准的网络数据。所谓认知社会结构（CSS），指的是这样的关系系统 $R_{i,j,k}$，其中 i 是关系的发出方，j 是关系的接收方，k 是对 i 和 j 之间关系进行评价的认知方（perceiver）（Krackhardt, 1987：113）。i，j，k 一般是同一个网络中的不同成员。CSS 的目标是把一系列不同的关系或者同一个网络被不同人评价后得到的不同认识片段（cognitive slices）合并成为一个组合的网络（pooled network）。

CSS 的具体任务如下。输入的文件是一个包含 k 个邻接矩阵（adjacency matrices）的多元矩阵数据文件，每个矩阵的形式为：$A(i, j)$ 嵌入三维矩阵

$A(i, j, k)$当中。这种形式有益于表征认知社会结构,其中 k 代表对 i 指向 j 的关系进行“认知”或评价的认知者。该程序主要使用两种方法对这种三维网络与一种二维网络 $A'(i, j)$进行比较。一种是根据 k 个矩阵计算对应元素总和,$A'(i, j)=A(i, j, k)$对 k 求总和。可以将这个矩阵根据一定标准进行二值化处理,从而产生一种“一致结构”(consensus structure)。另外一种方法是产生一种所谓具有组合性质的“局部汇集结构”(locally aggregated structure, LAS),方法是令 $A'(i, j)=A(i, j, i)+A(i, j, j)$。换句话说,在组合的矩阵中,既定单元格之值仅仅是对所涉及的两个行动者(而不是整个群体)的认知的函数(Krackhardt, 1987:115—117)。该矩阵也可以进行二值化处理。总之,CSS 的目的是根据一定的标准,将多个具有认知性的社会结构的矩阵组合成为一个矩阵。

3. 第三类:数据的展示及描述(Browse, Display, Describe)。

这部分比较简单,简述如下:

(1) Browse:数据矩阵浏览。

(2) Display:在计算机屏幕上展示 UCINET 数据库。

(3) Header browser/editor:打开一个数据矩阵,展示其维度以及每一行和每一列的标签。

(4) Describe(Obsolete):执行与 Header browser/editor 类似的功能,不过不经常使用。其具体的功能包括对数据进行描述(如描述数据的规模、维度、类型及标签等),并允许输入、输出、编辑标签,即将标签加入到行、列或者整个矩阵当中。

4. 第四类:数据的提取、移动、开包、合并—匹配分析(Filter/Extract, Remove, Unpack, Join, Match Datasets)。

(1) Filter/Extract:数据过滤或提取。具体包含如下六类操作。

Filter rows via SQL:利用 SQL(结构化查询语言,Structured Query Language)筛选矩阵的行。

Subgraph via attribute vector:通过属性向量从一个数据矩阵中提取出子矩阵。要执行此命令,需要有一个属性向量数据。该属性向量可以是数据矩阵的某一列,这需要研究者具体指定。

Main component:提取出与网络的弱成分对应的子图并加以保存。只要一个网络的任何两个点都可以通过长短不一的途径建立联系,就称这样的网络为

成分。所以，一个整体网络可能包含多个成分。有关强成分和弱成分的界定分类参见本书第六章。

Ego-network：提取出网络中某点的个体网成员及其间的关系，相当于该个体网络的所有成员构成的整体网（可包含该个体，也可不包含，在程序中可选）。

Submatrix：从数据矩阵中抽取出子矩阵（submatrix）。假设数据矩阵比较大，我们可能仅对其中的某些行动者之间的关系感兴趣，这时候就需要把这些行动者之间的关系网从总的关系网中"抽取出来"，此时就可利用该命令。具体采用如下所示的方法选出你需要的行动者及其间的关系：

3 to 10 或者 3-10

first 5

last 10

前两行指明（必须用行动者的代号，不能用标签）行动者的 id 号码为 3，4，5，6，7，8，9，10 的点。最后一项只有在总行动者数目确定的情况下才可用。这些约定可以混合在一起，如果希望把数据前两个点，第 4 点和第 5 点以及最后四个点选出来，则需要在"Which rows"和"Which columns"中键入（实际上在 Which rows 右侧键入需要选哪些点的时候，在 Which columns 右侧的框中会自动键入同样的内容）：

first 2，4，5，last 4

具体对话框如图 3.6 所示：

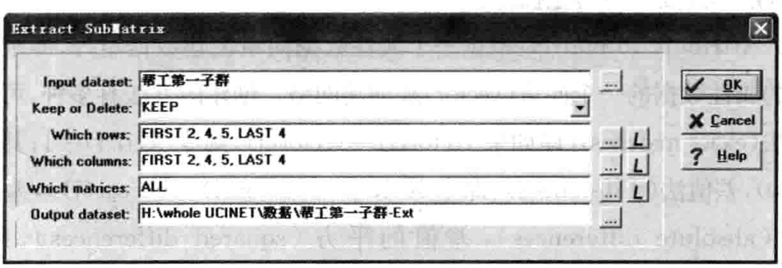


图 3.6　提取子矩阵的对话框

点击 OK 后，程序会自动给出 8 行 8 列的数据矩阵，对应的八个点是 1，2，4，5 和最后四个点。

Main Diagonal:抽取出主对角线的值。对于主对角线都是 0 的矩阵来说,该命令基本用不上。

(2) Remove:去除 UCINET 数据中的某些点。具体包括两种选项:一是移除孤立点(isolates),选择它即可移除网络中的孤立点,从而得到关联图;二是移除孤伶点(pendants),即与他者联系很少、度数只是 1 的点。对于非对称数据来说,该程序去掉的是那些出度(out-degree)为 1 的点。

(3) Unpack:对一个包含多种关系的矩阵数据进行开包处理,其结果是分成多个独立的矩阵,并加以保存,这样便于对单个矩阵进行分析。

(4) Join:与 Unpack 的功能相反,Join 命令把一系列 UCINET 数据合并成为一个数据文件,可根据行、列或整个矩阵进行合并。也可以将多个单独的矩阵或多元矩阵合并成为一个多元关系矩阵(multi-relational matrix)。当然,在合并的时候,各个矩阵的规模要匹配,即行数和列数要相等,否则不能合并。

(5) Match datasets:按照一定标准对网络数据进行匹配处理,保留两个数据中对应的元素一致的部分。

5. 第五类:数据的排序、置换、转置、匹配等(Sort, Permute, Transpose)。

(1) Sort alphabetically:按照字母顺序对一个网络中各个点进行排序。

(2) Sort by attribute:按照属性数据对一个网络中各个点进行排序。

(3) Permute:按照研究人员自己指定的顺序对行和列同时进行置换。

(4) Transpose:对数据矩阵进行转置处理。

6. 第六类:数据的其他操作。

(1) Attribute to matrix:根据一个属性数据向量创建数据矩阵,即根据包含多个点的属性数据的一个向量(vector)来创建网络。创建的方法有多种,可以是完全匹配法(exact matches)(即如果 vector(i)=vector(j),则令 X(i, j)=1;其他情况下等于 0),差值法(differences)(即令 X(i, j)=vector(i)−vector(j)),差值的绝对值法(absolute differences),差值的平方(squared differences),乘积法(product)或者数值之和(sums of the values)等。

(2) Affiliations(2-mode to 1-mode):将 2-模网络(隶属数据)转换为 1-模网络数据。分别根据行和列,可转换成两个 1-模网络。

(3) Matrix to vector:根据数据矩阵构造列向量(对角线的值可选)。

(4) Subgraphs from partitions:根据网络的分区情况提取出子图。

(5) Partitions to sets:根据行动者发生矩阵(actor incidence matrix)将一个分区指标向量(partition indicator vector)转换成一个群体,并且根据群体展示分区(display partition by groups)情况。一个分区指标的形式如下(k1, k2, …, ki …),其中 ki 代表将点 i 分派到群体 ki。因此,(1 1 2 1 2)将点 1, 2 和 4 分配到第一块,将 3 和 5 分配到第二块。所谓一个点—群体发生矩阵(group by vertex incidence matrix)指的是将点作为列,将群体作为行。在第 i 行,第 j 列的 1 表示行动者 j 是群体 i 的一个成员,其他情况下这些值是 0。

(6) Creat Node Sets:创建点集,即在比较两个向量或者一个向量与一个数字的基础上,创建一个群体指标向量(group indicator vector)。具体地说,在给定一个包含每个点的属性的向量或者取值向量以及一个临界值的情况下,该程序将选择出那些取值小于(或者大于)临界值的行动者。通常情况下,临界值本身也可以是一个向量,从而使得那些取值小于(或者大于)临界值向量的对应单元元素的行动者被选出来。例如,如果两个行动者的接近中心度都小于他们的度数中心度,则可以利用上述两个向量的方法选出这样的行动者。

(7) Reshape:对数据的规模加以重新组织,相当于从一个矩阵中提取一个新矩阵,当然,新矩阵的规模不同于初始矩阵。

(三) 数据转换菜单(Transform)

这个菜单包含一些把图和网络转换为其他类型的子菜单。它又分为三大类。

1. 第一类主要涉及数据的组合,即 Aggregate(包括 CSS)。具体包括两个命令 Block 和 Collapse。

Block:把一个数据中的各个点进行分块,计算块的密度(block densities)、块中每一个元素的均值、最大值、最小值等。这个命令与 Networks→Properties→Density 命令基本相同,后者仅仅计算出块的密度。

Collapse:压缩,即将一个矩阵的多行或多列组合在一起。该命令允许对各个行和列分别对待,允许使用者选择如何计算复合值(combined values)。例如,利用 Extract 命令,将 UCINET 自带的数据 KNOKBUR(该数据包含两个矩阵)分成两个子矩阵 KNOKI 和 KNOKM,其中 KNOKI 数据如表 3.8 所示:

表 3.8　KNOKI 信息交换数据

	1	2	3	4	5	6	7	8	9	10
	CO	CO	ED	IN	MA	WR	NE	UW	WE	WE
	UN	MM	UC	DU	YR	O	WS	AY	LF	ST
	--	--	--	--	--	--	--	--	--	--
1	0	1	0	0	1	0	1	0	1	0
2	1	0	1	1	1	0	1	1	1	0
3	0	1	0	1	1	1	1	0	0	1
4	1	1	0	0	1	0	1	0	0	0
5	1	1	1	1	0	0	1	1	1	1
6	0	0	1	0	0	0	1	0	1	0
7	0	1	0	1	1	0	0	0	0	0
8	1	1	0	1	1	0	1	0	1	0
9	0	1	0	0	1	0	1	0	0	0
10	1	1	1	0	1	0	1	0	0	0

如果由于某种原因，希望将 KNOKI 数据的第 1 列到第 4 列合并加在一起，则需要依次点击 Transform→Aggregate(includes CSS) →Collapse，在出现的对话框"Input dataset"中选出 KNOKI，在"Enter instructions for collapsing"中键入：

Cols 1 to 4

并且在对话框中选择合并的标准为"sum"。点击 OK 后，即可得到如下合并的结果：

	1	2	3	4	5	6	7
	B1	MAY	WRO	NEW	UWA	WEL	WES
	---	---	---	---	---	---	---
1	1	1	0	1	0	1	0
2	3	1	0	1	1	1	0
3	2	1	1	1	0	0	1
4	2	1	0	1	0	0	0
5	4	0	0	1	1	1	1
6	1	0	0	1	0	1	0
7	2	1	0	0	0	0	0
8	3	1	0	1	0	1	0
9	1	1	0	1	0	0	0
10	3	1	0	1	0	0	0

可见，合并后第一列的数据就是矩阵合并前第 1 列到第 4 列对应元素之和。

2. 第二类包含多个命令，主要针对矩阵的元素进行处理，具体内容包括如下几类。

(1) Dichotomize：二值化处理。按照一定标准，将多值数据矩阵进行二值化处理。

(2) Dichotomize interactive：二值化处理。该命令给出一些具体的计算指标，包括数据中每个值的 z-值，最大值、最小值等。

(3) Symmetrize：按照一定标准，将数据矩阵进行对称化处理。

(4) Transitivize：按照一定标准，将数据矩阵进行传递化处理。具体依据的标准可以有三类：

第一类：如果在初始的数据矩阵 X 中，$x_{ij}=1$(即第 i 行第 j 列的取值为 1)，$x_{jk}=1$，那么在经过传递化处理后得到的新矩阵 Y 中，其元素的取值为 $y_{ij}=1$，$y_{jk}=1$，$y_{ik}=1$。

第二类：如果在矩阵 X 中，$x_{ij}=1$，$x_{jk}=1$，那么在新矩阵 Y 中，$y_{ik}=1$，$y_{jk}=0$，$y_{ik}=0$。

第三类：如果在矩阵 X 中，$x_{ij}=1$，$x_{jk}=1$，$x_{ik}=0$，那么在新矩阵 Y 中，$y_{ij}=1$，$y_{jk}=1$，$y_{ik}=1$。

(5) Transpose：将矩阵进行转置处理。

(6) Normalize：按照一定标准，将矩阵的行、列或者整个矩阵进行标准化处理。

(7) Match Marginals：按照边缘值对数据矩阵进行标准化处理。

(8) Recode：对矩阵重新编码。

(9) Reverse：取反值，即用矩阵中最大值与最小值之和减去矩阵的每一个值。这相当于利用某种线性转换，将相似性(similarity)数据转换成距离(distance)数据，或者距离数据转换成相似性数据。对于 0-1 矩阵来说，该程序相当于将 1 改成 0，将 0 改成 1。

(10) Diagonal：对角线命令，可用该命令抽出矩阵对角线的值，或者改变矩阵对角线的值(如将全部对角线的值 0 都改为 1)，并加以保存。

(11) Double：按照一定标准，对一个数据的各列进行双倍处理，处理后得到的矩阵的列数是原矩阵列数的两倍。

(12) Rewire：按照某种标准重新处理矩阵，达到某种优化。可选的标准有八个：聚类系数(clustering coefficient)最大化、最小化，距离最大化、最小化，聚类系数最大化并且距离最大化，聚类系数最大化并且距离最小化(在这种情况下，整个网络就是一个小世界)，聚类系数最小化并且距离最大化，聚类系数最小

化并且距离最小化。要注意,这种操作改变了矩阵的信息,产生的是一个满足研究者所指定条件的新矩阵。

(13) Matrix Operations:矩阵算法,针对一个或多个矩阵进行各种计算。主要包括矩阵内算法(Within Dataset)和矩阵间算法(Between Datasets)。所谓矩阵内算法指的是只针对一个矩阵进行的计算,包括矩阵内的汇总(Aggregations)和矩阵中每个元素的转换(Cellwise tranformations)。前者会计算出矩阵每一行(或列)的总和,或者计算矩阵各个值的总和;后者会针对矩阵中的每一个值进行转换,例如计算绝对值、对数值、平方值等。

矩阵间算法指的是针对两个矩阵进行的计算,包括统计汇总(Statistical Summary)和布尔代数组合(Boolean combinations)。可执行两个矩阵的加、减、平均、元素乘积(elementary multiplication)以及布尔代数计算等。有关矩阵运算的具体操作,参见本节第三部分。

(14) Make Interaction Term for Regression:计算两个矩阵的对应元素之积(multiply),不是矩阵算法中的两个矩阵之积(production)。这相当于计算两个矩阵变量的交互项,可将计算结果用于矩阵的回归分析。在计算之前可对两个矩阵进行对中处理(mean-center)或标准化处理。

3. 第三类命令包含五个,主要针对矩阵进行其他类型的转换,具体内容如下。

(1) Graph Theoretic:一些图论命令。具体包括如下子命令。

Bipartite:把一个二部图(bipartite graph)的发生阵(incidence matrix)(即 2-模数据矩阵)转换为一个 1-模的邻接阵。如果发生阵有 n 行 m 列,那么作为转换结果的邻接矩阵就是一个有 m+n 行和列的方阵。

Make Incidence Matrix:制作一个发生阵。把一个邻接矩阵转换为一个长方形的点—线指标矩阵(node-by-line indicator matrix)。

Linegraph:线图。针对有向图,可用它来创建一个线图,其中的点对应于初始图中的各条线。

Multigraph:多图。把一个多值图(valued graph)转换为一系列二值邻接矩阵(binary adjacency matrices)。假设一个多值矩阵的元素取值有 0, 1, 2, 3,那么该程序会计算出三个矩阵。第一个矩阵将所有的元素取值为“1”的值变为“1”,其余值都变为“0”;第二个矩阵将所有取值为“2”的值变为“1”,其余值都变为“0”;第三个矩阵将所有取值为“3”的值变为“1”,其余之值都变为“0”。

Multiplex:制作多丛图。可以从一个多元关系图(multirelational graph)中

构建一个多丛图(multiplex graph),即计算出每一对点之间共有多少种关系。

Semigroup:制作半群图。该程序根据图、有向图或者多元图来构造半群(semigroup)。所谓一个网络的半群(semigroup of a network)指的是对其全部组合关系的代数表征。具体地说,假设一个多元关系中的一系列二值关系矩阵为 R_1, R_2, …, R_n,那么全部可能的一对矩阵的布尔代数乘积(Boolean products of pairs of matrices)给出的则是所有可能的长度为 2 的关系。我们继续计算长度为 3 的矩阵之积,直到不出现新的矩阵。按照这种布尔代数乘积方式构造的全部矩阵就构成一个半群(semigroup)。该程序可发现半群的全部成员。半群分析一般很少用到。

Union:将多个矩阵合并成一个数据矩阵,即将匹配的两个或多个矩阵对应元素相加。

Intersection:取多个关系矩阵的合集,即将独立的两个或多个矩阵依次排列,形成一个包含多个数据矩阵的文件。如果两个矩阵是匹配的,则形成的合集数据其实就是这两个数据矩阵的依次排列。如果两个矩阵不匹配(即行数和列数不等),那么计算得到的合集数据也是两个矩阵,不过其行数是这两个数据矩阵的行数之和,列数是这两个数据矩阵的列数之和。

(2) Time Stack:多时段网络的合并。将在不同时间段得到的同一群行动者之间的关系矩阵合并在一起。

(3) Egotize:整体网成员的个体网分析。制作整体网中每个成员的个体网(可含成员本人,也可不含)。如果整体网的规模为 n×n,执行该命令就得到 n 个新的 n×n 矩阵,这 n 个矩阵依次对应整体网中 n 个成员的“个体网成员”之间的关系,非个体网成员之间的关系都是 0。

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	1.00	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00	1.00	0.00
3	1.00	1.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	1.00	1.00
5	0.00	1.00	0.00	1.00	0.00	1.00
6	0.00	0.00	0.00	1.00	1.00	0.00

例如,在上面这个 6 点矩阵中,点 1 的个体网成员有 2, 3, 5。而在这个整体网络中,点 2, 3, 5 这三个点之间的关系如下:点 2 和点 3 相互有关系,点 2 和点 5 也相互有关系。因此,单独提取出 2, 3, 5 的关系,就会得到下面的矩阵。这就是"点 1 的个体网矩阵"。

Matrix: 1

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	1.00	0.00	1.00	0.00
3	0.00	1.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	1.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00

同理,点 2 的个体网成员有 1, 3, 5。在这个整体网络中,我们现在分析点 1, 3, 5 的关系。可以看到,点 1 和点 3 有相互的关系,点 1 有指向点 5 的关系,但是点 5 没有指向点 1 的关系,点 3 和点 5 无关系。单独提取出 2, 3, 5 的关系,就会得到下面这个"点 2 的个体网矩阵"。

Matrix: 2

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	0.00	1.00	0.00	1.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00
3	1.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00

由于该网络的规模为 6,因此,除了上述四个矩阵之外,该程序会再给出其他四个矩阵。这里从略。

(4) Build Ties Change Matrices:建构随时间变化的数据矩阵。对于至少在

两个时间段建立的网络 T1、T2 等来说，可以有三种方式建构关系变化的矩阵。第一种是计算两个矩阵之差，即计算 T2－T1，计算结果是 Y12(i, j)＝T2(i, j)－T1(i, j)。第二种是仅计算两个矩阵之增量，即如果 T2≥T1，则计算 Y＝T2－T1，其余的 Y 值都设定为缺失值 0。第三种是计算关系的形成，即如果 T2≥T1，则令 Y＝T2－T1＝1，其余的 Y 值都设定为缺失值或 0。

Matrix: T1

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	0.00	0.00	0.00	3.00	0.00
2	0.00	0.00	0.00	0.00	3.00	2.00
3	0.00	2.00	0.00	0.00	0.00	0.00
4	0.00	0.00	2.00	0.00	3.00	0.00
5	2.00	3.00	0.00	1.00	0.00	0.00
6	0.00	3.00	0.00	0.00	0.00	0.00

Matrix: T2

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	1.00	0.00	0.00	3.00	0.00
2	0.00	0.00	0.00	0.00	3.00	1.00
3	0.00	1.00	0.00	0.00	0.00	0.00
4	0.00	0.00	2.00	0.00	3.00	0.00
5	0.00	3.00	0.00	2.00	0.00	1.00
6	0.00	3.00	0.00	0.00	2.00	0.00

如果按照第一种计算方法，会得到如下结果。可见，该结果就是上述两个矩阵对应元素之差 T2－T1。该结果记载了前后两个矩阵的变化细节，包括增量、减量等。

Matrix: T1T2

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	1.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	-1.00
3	0.00	-1.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	-2.00	0.00	0.00	1.00	0.00	1.00
6	0.00	0.00	0.00	0.00	2.00	0.00

按照第一种方法计算结果

如果分别按照第二种和第三种计算方法，会得到如下两个结果。第二种结果仅记录了“增量”部分，没有记录“减量”。第三种结果仅记载了是否有增量，没有记录“减量”，但用缺失值记录了前后都有关系并不变的情况。

```
T1T2            T1T2

  123456          123456
  ------          ------
 1010000         10100 0
 2000000         20000 0
 3000000         3000000
 4000000         40000 0
 5000101         50 0101
 6000020         60 0010
```

按照第二种方法计算结果　　按照第三种方法计算结果

(5) Give Non-responders Missing Rows:将一个矩阵中无回答者视为缺失值。要执行此计算,需要一个邻接矩阵和一个匹配的属性行向量或列向量,属性向量中的1表示对应的数据有效,0表示无效。该程序根据该向量中的1和0对邻接矩阵进行重新展示,即不再显示缺失值。

(四) 计算工具菜单(Tools)

UCINET 提供的计算工具出现在"Tools"菜单下,包含经常使用的"技术工具"。尽管其中的某些技术也被一些非网络者使用,如多维量表分析(multidimensional scaling)和聚类分析技术(cluster analysis)等,但是在 UCINET 中,它们分析的对象是关系数据。这些工具命令也分为如下三大类。

1. 第一类命令包含如下重要命令。

(1) Profit:是 Property Fiting 的缩写,执行拟合性质的分析。它对一系列坐标值(co-ordinates)和一系列连续属性值进行拟合性质分析。假定有一系列坐标值(如从多维量表分析得到的 X 坐标和 Y 坐标值)和一个连续的属性变量,该程序会利用多元回归分析发现坐标值对属性变量的最佳拟合情况,并用一个二维图展示。另外,它用一种置换检验方法(permutation test)给出拟合情况的显著性水平。该算法常用于根据行动者属性测度来检验有关多维图内在维度的假设。

因此,该程序需要导入的坐标数据文件包含二维度的坐标值,即第一列表示 x 坐标,第二列表示 y 坐标,数据类型因而是多值的 n×2 矩阵。需要导入的属性数据则应包含行动者的属性,其中行动者 i 在第 j 类属性上的值应该体现在第 i 行第 j 列中,数据类型是多值数据。

（2）Consensus analysis：分析多个答题者在回答问题方面的一致性。参见本书第十章。

（3）Cluster analysis：对矩阵数据进行聚类分析，包括层次聚类、组合式最优化分析、聚类充分度分析。

（4）Scaling/Decomposition：量表及数据分解。它又包括六小类子命令。

Metric-MDS：针对相近性矩阵（proximity matrix）进行量纲式的多维量表分析（multidimensional scaling）。

Non-metric MDS：针对相近性矩阵进行非量纲式多维量表分析。一个相近性矩阵可以是相似性矩阵（similarities）（即其取值越大，表示行动者之间的关系越近），也可以是相异矩阵（dissimilarities）（即其取值越大，表示行动者之间的关系越远），一般情况下是距离矩阵。该程序会在 k-维空间（k-dimensional space）发现这样的一系列点，即这些点之间的欧氏距离（Euclidean distances）尽可能接近于输入的相近性矩阵。输入的矩阵必须是一个对称的方阵。

Factor Analysis：因子分析。

Correspondence Analysis：对 2-模网络进行对应分析。

Eigenvectors of nonsymmetric matrix：计算非对称方阵的特征向量和特征值。

SVD（Singular Value Decomposition）：针对 2-模网络数据进行奇异值分解。

2. 第二类命令包含如下内容。

（1）Correlate columns across datasets：计算两个 n×n 矩阵各列之间的相关系数矩阵，得到的是一个 2n×2n 矩阵。

（2）Similarities（e.g.，correlations）：相似性分析。根据一定标准（如相关系数、方差、雅可比系数等）计算出矩阵各行或各列之间的相似性系数矩阵。

（3）Dissimilarities & Distances：相异性分析和距离分析。根据一定标准（如欧氏距离、马氏距离、雅可比距离等）计算出矩阵各行或各列之间的相异性系数矩阵。

（4）Univariate statistics：对一个矩阵中的值进行单变量统计分析。可以对一个矩阵的每行（或者每列，或者整个矩阵）计算其均值、标准差、方差、最小值、最大值和总观察值等值。

（5）Frequencies：对行或列进行频次分析。

（6）Count combinations：针对整个网络，计算其各行的模式发生的频次

(frequency of pattern of rows)。简而言之,该程序将计算每一行的数字完全一致的行有多少,然后按照最常发生的行到最不经常发生的行进行排列。

(7) Testing hypothesis:假设检验,包括三类。

第一类:Node-level,点层次的假设检验。包括回归分析(regression)、方差分析(anova)和 t 检验(t-test)。

第二类:Mixed Diadic/Nodal,点—关系混合层次的假设检验。

第三类:Diadic QAP,关系—关系层次的假设检验。它又包括如下几类:

QAPcorrelation:矩阵相关分析。

QAPcorrelation(old):旧版的矩阵相关分析。

QAPrelational Crosstabs:矩阵关系的列联表分析。

MR-QAP linearregression:基于多元回归二次指派程序(Multiple Regression Quadratic Assignment Procedures)的矩阵线性回归分析。

LR-QAP logisticregression:QAP 矩阵的 logistic 回归(logistic regression)分析。

(8) Command line/Matrix Algebra:矩阵代数分析。这个命令很有用,可利用它来执行矩阵之积、逆矩阵、特征值等运算。在后面的讲解中会涉及其中一些命令的使用。

3. 第三类命令包含三个画图命令,分别如下。

(1) Scatterplot:散点图。

(2) Dendrogram:树状图。

(3) Tree Diagram:树形图。

(五) 网络分析菜单(Network)

在 UCINET 中,"Network"菜单下附带的则是一些基本的"网络分析技术",如中心性分析、核心—边缘分析、子群分析等技术。这些命令是社会网络分析的核心,分为五大类,如图 3.7 所示。

1. 第一类命令包含四个子命令,分别如下。

(1) Cohesion:凝聚性分析。可计算的凝聚力指标有十几种,如整体网的密度、互惠性、同质性、各点之间的距离矩阵、聚类系数,各点之间的可达性(reachability)、传递性、E-I 指数等。

(2) Regions:区域分析。计算并发现"成分"(components),包括强成分和弱成分、双成分(Bi-components)和 k-核(k-Cores)。

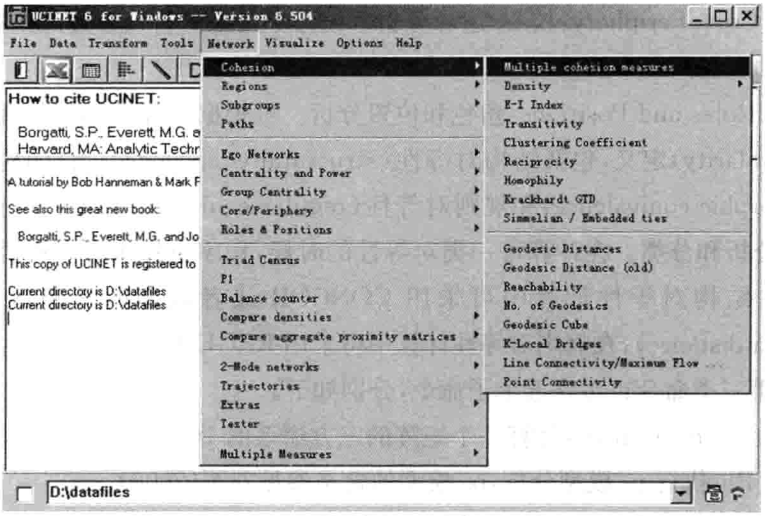


图 3.7 Network 命令包含的子命令

(3) Subgroups:子图分析。可用来计算各种类型的凝聚子群,包括三小类共九个子命令。第一校类包括派系分析(cliques)、n-派系(n-cliques)、n-宗派(n-clans)、k-丛(k-plexes)分析等。第二小类包含 Lambda 集合(Lambda sets)、指定数目的派系(factions)和计算出满足 G-N 算法(Girvan-Newman algorithm)的凝聚子群,具体参见其中的 Help。最后一小类包括根据弱传递性发现互斥的凝聚子群(f-groups),根据马尔科夫聚类算法(Markov clustering)计算出不重复的凝聚子群。各个子命令的具体含义比较复杂,这里不介绍了。

(4) Paths:路径分析,分析各个点之间存在的路径。

2. 第二类命令包含五个子命令,分别如下。

(1) Ego Networks:个体网分析,又包含九个子命令。该选项会计算出网络中每个行动者的一系列个体网指标,如个体网的基本指标、结构洞、中间人角色、个体网的相似性等。

(2) Centrality and Power:中心性和权力分析,包括六类指令,每一类又包括一些命令。可计算各种点的中心度指数,包括度数中心度(degree)、接近中心度(closeness)、中间中心度(betweenness)、流量中间中心度(flow betweenness)、信息中心度(information centrality)、特征向量中心度(eigenvector centrality)、权力指数(power)和 Katz and Hubbell 测度等。还可以计算出 2-模网络的中心性指数。

(3) Group Centrality:群体的中心性分析。

(4) Core/Periphery:核心/边缘分析。该命令用来分析核心边缘结构,确定每个行动者在网络中的位置。

(5) Roles and Positions:角色和位置分析。可根据几种结构相似性(structural similarity)定义,包括结构对等性(structural equivalence)、自同构对等性(automorphic equivalence)和规则对等性(regular equivalence)等,对行动者进行聚类分析和分类。在计算每一类对等性的时候,程序也提供了多种算法的选择(如在结构对等性计算中可采用 CONCOR 或者截面欧氏距离(profile euclidean distance);在规则对等性计算中可采用 REGE 和 CATREGE。

3. 第三类命令也包含五个子命令,分别如下。

(1) Triadic Census:计算一个矩阵的三方谱系的 16 个同构类。

(2) P1:执行 p_1 模型分析,p_1 模型的含义参见刘军(2006)。

(3) Balance Counter:符号三方关系计算。已知一个集合的行动者之间的两类有向关系,一个关系是正的,另外一个是负的,还已知行动者的分区情况,那么该程序会计算出使用者所指定的三方组(user specified triads)。这样的三方关系共有 216 种可能。该程序会计算出每一类三方组,允许使用者指定哪一类。对于每一类三方关系(tripple)i, j, k 来讲,我们要考察 i 到 j, j 到 k 和 i 到 k 的关系,即 iRj, jRk 和 iRk,计算 R 是正、负或无的关系的案例数,同时也考察每一对点(i, j),(j, k)和(i, k)处于相同或不同分区中的数目。最后分别根据整个网络和行动者层次给出计算结果。

(4) Compare Densities:计算两个匹配的矩阵密度之差,并给出显著性检验。也可以计算一个矩阵的密度与理论上的密度之差,并给出显著性检验。参见本书第十二章最后一节。

(5) Compare Aggregate Proximity Matrices:对多个组合的相似性矩阵进行比较。

4. 第四类命令包含四个子命令,分别如下。

(1) 2-mode networks:计算 2-模网络的凝聚性、分派性、中心性以及核心/边缘分析。

(2) Trajectories:针对一个 2-模职业轨迹矩阵(2-mode career trajectories matrix),计算其中间中心度和一些基本测度。所谓“职业轨迹矩阵”是这样一个 2-模矩阵,其中的行代表行动者,列代表职位或工作。第 i 行第 j 列的元素给出的是行动者求得职位 j 的时间。时间是定序数据,它给出的是第 i 个人在其特定

的职业轨迹中遵循的顺序。该矩阵必须包含一列用来表明退休时间，该列通常是最后一列。该程序将针对每一个位置考察职业数据中累积的职位，并且给出各类职位的角色和中间中心度信息。

(3) Extras：有关网络分析的其他指标，具体包括十几种。

(4) Tester：检验数据的规模。相当于展示数据的具体行值和列值。

5. 最后一类命令包含一个子命令。

Multiple Measures：同时展示一个矩阵的多个凝聚力指标或矩阵中每个点的各种中心度指标。

(六) 可视操作菜单(Visualize)

包括三个选项：NetDraw，Mage 和 Pajek。选择其中的一个命令，会调出相应的画图程序，对数据矩阵进行画图处理。

(七) 选项命令菜单(Options)

选项菜单包含一些可供选择的命令。这些选项可以改变 UCINET 6 系统参数的默认设置值，其中重要的命令包括如下。

Data Checking：数据检查。

Display Graphical Dendrograms：展示图形。

Display Full Pathnames：展示完整的路径名。

Smart Default Name：优化缺省名。

Unicode Labels：统一码标签。

UCINET File Type：设定 UCINET 的文件类型。

Dicimal Place：设定小数点的位置，一般确定在小数点后三位。

Width of columns：设定列的宽度。

Page Size：改变页码的规模。

Scratch Folder：改变数据分析使用的文件夹。

Output Folder：修改输出数据所在的文件夹。

Helper Application：锁定各类帮助命令所在的文件夹。

Repeat command：重复刚刚使用的命令。

(八) 帮助菜单

帮助子菜单包含注册信息(Register)、帮助主题(Help Topics)等内容，其中帮助主题最重要。点击它可以找到很多专业术语的含义以及在 UCINET 中的分析步骤。虽然 UCINET 的程序在不断更新，但是其帮助主题中的很多内容都

没有更新。有关程序的含义需要读者自行查找资料进行分析。

(九) 快捷方式

UCINET 程序是菜单式的,便于应用。有时候可用一些快捷键。某些菜单(如 Data→Display)已经被赋予热键(hot-keys),因而可以迅速调出来,用不着通过一些菜单。假设没有菜单可选,直接键入 Ctrl+D 就会立即激活 Data→Display 路径。

所有的快捷键都用 Control 或者 Alt 加上字母的组合。如下所示:

Ctrl+B	Data→Header/Browser
Ctrl+D	Data→Display
Ctrl+F	File→Change Default Folder
Ctrl+M	Network→Centrality and Power→Multiple Measures
Ctrl+S	Data→UCINET Spreasheet
Alt+X	File→Exit

三、UCINET 中矩阵运算的句法

在 UCINET 中,矩阵的各种运算都可利用一定的矩阵算法进行。总的来说,这些算法都具有如下形式:output matrix=function(arguments)。具体计算的路径为:在 UCINET 中,选择 Tools→Matrix Algebra 这条路径,键入相应的命令,回车即可得到计算结果。前一节因篇幅所限,不能介绍矩阵运算的细节,下面加以介绍。针对矩阵的计算主要有如下四类:

(一) 一元操作(Unitary Operations)

它只针对矩阵本身进行操作。例如,

i=id(100)(生成一个规模为 100 的单位矩阵)

mat(〈real〉, [〈nr〉], [〈nc〉], [〈n1〉])(把一个数字转变为一个矩阵,或者产生一个常数矩阵)。如果行数〈nr〉、列数〈nc〉和层次数〈nl〉没有被指定,那么该程序将生成一个 1 行 1 列的常数矩阵,它只包含一个数字〈real〉。参数〈n1〉指定矩阵的层次(或者数目)。如果指定了〈n1〉,那么必须也指定行数〈nr〉和列数〈nc〉。例如,

junk=mat(3.92)(生成一个 1 行 1 列的矩阵,矩阵元素取值就是 3.92,命名为 junk)

junk=mat(4, 10, 10)(生成一个 10 行 10 列的矩阵,其中的值都是 4,并将

该矩阵命名为 junk)

junk＝mat(4, 10, 10, 2)(生成两个 10 行 10 列的矩阵,其中的值都是 4,并将该多元矩阵命名为 junk)

(二) 二元操作(Binary Operations)

该类计算至少针对两个矩阵进行操作。其操作格式为:输出矩阵＝运算法则(两个或多个矩阵)。下面介绍几种操作句法。

(1) 假设 x 是当前文件夹中已知的一个数据,如果计算 x 的逆矩阵,需要用到 inverse 命令。令 y 代表即将生成的 x 的逆矩阵,则求 x 的逆矩阵的命令为:

y＝inverse(x)

(2) 矩阵加法运算指令为:〈matrix〉＝add(〈matrix1〉, 〈matrix2〉, ...)

(3) 矩阵相减的句法为(Subtract)。命令为 sub(〈mat1〉, 〈mat2〉, ...)(其含义为前一个矩阵和后一个矩阵的对应元素相减。例如,c＝sub(a, b))。

(4) 可使用复合命令,例如,

y＝inverse(transpose(inf))(计算矩阵 inf 的转置矩阵的逆矩阵,并保存为矩阵 y)。

(5) 还可以用更复杂的语句计算所需的矩阵,如:

b＝prod(inv(prod(transp(x), x)), prod(transp(x), y))。

但是这样容易出现错误,最好分解命令,可以改变为:

xt＝transp(x)

xtx＝prod(xt, x)(计算矩阵之积)

xty＝prod(xt, y)

b＝prod(inv(xtx), xty)

其他句法举例如下。

(6) 布尔代数积(Boolean Product),其句法为:

bprod(〈mat1〉, 〈mat2〉)(计算两个二值矩阵的布尔代数积)。例如,

junk＝bprod(business, marriage)

(7) 乘积(Multiply),句法为 mul(〈mat1〉, 〈mat2〉, ...)(计算两个或者多个矩阵对应值之积)。例如,c＝mul(a, b)将计算出矩阵 a 和 b 的积,结果矩阵命名为 c。

(8) 乘积(Product),句法为 prod(〈mat1〉, 〈mat2〉, ...)。它是矩阵之积,而不是矩阵的对应元素之积,因而与 MULTIPLY 算法不同。例如,

buskin＝prod(business，marriage)。该句法计算商业关系矩阵和婚姻关系矩阵的乘积，结果为 buskin 矩阵。

数据也可以使用其完整的路径名，如 a：tdavis＝transpose(c：\UCINET 6\data\davis)。该语句表示，对 C 盘中的 UCINET 6 文件夹中的 data 文件夹中的 davis 数据矩阵求转置，结果放在 A 盘中，并命名为 tdavis。

需要补充的是，计算两个矩阵之积的含义是什么？简单地说，由于我们往往约定矩阵的含义代表的是“发送—接收”关系，因此，两个矩阵之积代表“发送—接收关系”再乘以“发送—接收关系”，它等于新的“发送—接收关系”。实际上给出的是两步途径距离的情况(即发送者与“接受者的接受者”之间的关系情况)

如果是计算矩阵的幂次，道理类似。总的来说，矩阵 A^p 中各项给出的是从点 i 到点 j 的长度为 p 的线路(walk)总数(Wasserman & Faust，1994：159)。

(三) 矩阵内运算(Inner Products)

矩阵内运算的一些句法主要针对一个矩阵的各种性质进行计算，例如，矩阵中关系总数，行总和等。举例如下。

nties＝tot(davis)(该算法将计算出 davis 这个矩阵中的总关系数，并命名为 nties)

tdavis＝transp(davis)(该算法将计算出 davis 这个矩阵中的转置矩阵，并命名为 tdavis)

(四) 程序(Procedures)

(1) 展示(Display)。对应的句法为 disp 〈mnat〉 or dsp 〈mat〉。它将在屏幕上展示出〈mat〉的全部元素。此类例子又如：

dsp C：\UCINET\data\padgett

展示(display，简写为 dsp)C 盘 UCINET 文件夹中 data 中的 padgett 这个矩阵。

(2) inv(camp92)。该命令将针对方阵 camp92 进行逆矩阵分析(inverse)，该命令展示出 camp92 这个矩阵的逆矩阵。

(3) ginv(〈mat〉)将计算出一个矩阵的一般化的逆矩阵。

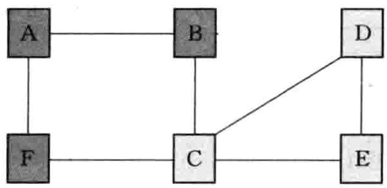
把 dsp 和 ginv 或者 inv 结合，可以输入如下语句：

dsp ginv(davis)(展示 davis 这个矩阵的一般化的逆矩阵)。

dsp inv(camp92)(展示 camp92 这个矩阵的逆矩阵)。

（五）矩阵计算例示

假设我们研究六个城市人和农村人之间的关系，通过调查，得到如下关系图。假设 A、B、F 来自“城市”（用深色标出），C、D、E 来自“农村”（用浅色标出）。现在的问题是，如何计算出市民之间的关系数、村民之间的关系数以及市民和村民之间的关系数目？我们可以通过矩阵的操作来计算这些指标。



	X	X^2	X^3	距离矩阵
A	0 1 0 0 0 1	2 0 2 0 0 0	0 4 0 2 2 4	. 1 2 3 3 1
B	1 0 1 0 0 0	0 2 0 1 1 2	4 0 6 1 1 0	1 . 1 2 2 2
C	0 1 0 1 1 1	2 0 4 1 1 0	0 6 2 5 5 6	2 1 . 1 1 1
D	0 0 1 0 1 0	0 1 1 2 1 1	2 1 5 2 3 1	3 2 1 . 1 2
E	0 0 1 1 0 0	0 1 1 1 2 1	2 1 5 3 2 1	3 2 1 1 . 2
F	1 0 1 0 0 0	0 2 0 1 1 2	4 0 6 1 1 0	1 2 1 2 2 .

图 3.8　带有城乡属性的 6 点图及其邻接矩阵、距离矩阵

在 UCINET 中，首先将 X 数据输入，然后点击 Tools→Command Line/Matrix Algebra，在出现的对话框的“Type command below”处键入“X2＝prod(X, X)”，回车后即可计算出 X 的平方矩阵，同理，键入“X3＝prod(X2, X)”，回车后即可计算出 X 的三次方矩阵。为了计算距离矩阵，需要点击 Network→Cohesion→Geodesic Distance(old)，在“Input dataset”中选择 X 矩阵进行分析，即可计算出距离矩阵。

对多个矩阵进行运算，可以使我们注意到关系混合的模式（mixing patterns），即多类关系之间的关联模式。仅仅把邻接矩阵乘以一个类别指标即可看到关系的模式。例如，对于之前的关系图来说，如果 A、B、F 来自“城市”，C、D、E 来自“农村”，那么在 UCINET 中首先构造“来源矩阵”，其取值如下所示。然后计算矩阵 X 与“来源矩阵”之积。

在 UCINET 中，输入“来源矩阵”，并命名为“source”，然后在“Type command below”处键入“xsource＝prod(x, source)”，回车后即可计算出“X”与

“来源矩阵”的乘积。从计算结果可以看出，对于市民 A 来说，他与两个城里人有联系，与农村人没有任何联系；而对于 B 来说，他与一个城里人有联系，与一个农村人有联系；对于村民 C 来说，他与两个城里人、两个农村人都有联系；对于 D 来说，他与城里人没有任何联系，但是与两个农村人有联系。读者也可根据“X・来源矩阵”来解释 E 和 F 的关系模式。

$$
\begin{array}{c|c|}
\text{点} & X \\
\hline
A & 0\,1\,0\,0\,0\,1 \\
B & 1\,0\,1\,0\,0\,0 \\
C & 0\,1\,0\,1\,1\,1 \\
D & 0\,0\,1\,0\,1\,0 \\
E & 0\,0\,1\,1\,0\,0 \\
F & 1\,0\,1\,0\,0\,0 \\
\hline
\end{array}
\times
\begin{array}{|cc|}
\multicolumn{2}{c}{\text{来源矩阵}} \\
\hline
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
1 & 0 \\
\hline
\text{城} & \text{乡} \\
\hline
\end{array}
=
\begin{array}{|c|cc|}
\multicolumn{3}{c}{X\cdot\text{来源矩阵}} \\
\hline
A & 2 & 0 \\
B & 1 & 1 \\
C & 2 & 2 \\
D & 0 & 2 \\
E & 0 & 2 \\
F & 1 & 1 \\
\hline
 & \text{城} & \text{乡} \\
\hline
\end{array}
$$

图 3.9　城乡关系模式的计算

如果我们继续研究，在由这六个点构成的网络中，城里人之间的联系、城乡的联系以及农村人之间的联系有多少，则可以进行如下计算，即用“来源矩阵的转置矩阵”乘以“邻接矩阵 X 与来源矩阵之积”，这需要用到 UCINET 中的“转置”(transpose)命令和乘积命令，如下所示：

$$
\text{来源矩阵}^{T}\cdot(X\cdot\text{来源矩阵}) =
\begin{array}{c|cc}
 & \text{城} & \text{乡} \\
\text{城} & 4 & 2 \\
\text{乡} & 2 & 6
\end{array}
$$

计算结果表明，城里人之间的联系有 4 个，城乡之间的联系有 4 个(从城到乡的 2 个加上从乡到城的 2 个)，农村人之间的联系则有 6 个。考虑到这是无向关系矩阵，因此，每个值都要除以 2。也就是说，城里人之间的联系有 2 个，城乡之间的联系有 2 个，农村人之间的联系则有 3 个。

上述计算仅仅针对的是本例。实际上，读者完全可以根据自己收集到的案例数据进行相应的计算。例如，通过进行类似的计算，我们可以计算出“社会资本高”的人之间的联系有多少，“社会资本低”的人之间的联系有多少，以及“社会资本高”的人和“社会资本低”的人之间的联系有多少。这种计算在某些情况下是很有意义的。

第二节　UCINET 数据的输入

由于 UCINET 可分析的数据类型多，因此 UCINET 数据的输入也有多种。一般情况下，社会网数据都是通过社会调查，以问卷或者数据表的形式获得的。调查得到数据之后，必须在电脑上输入这些数据。输入的方式多种多样，可以用 Excel 或常见的文本编辑器输入，也可利用 UCINET 本身的数据表程序（Matrix spreadsheet）输入。如果输入的数据是 Excel 类型或文本类型的数据，可以在 UCINET 中利用 Data→Import Excel 或者 Data→Import Text File 命令把这些数据转换为一个 UCINET 6 数据文件。

Import 可以根据多种方式对 ASCII 码文件进行处理。就最一般情况而言，它可以导入三类数据文件：初始数据（Raw）、Excel 数据和数据语言数据（Data Language，DL）。初始数据矩阵仅仅包含一些数字，如一个回答者—变量矩阵（respondent-by-variable matrix）记录的是一些回答者在回答问题时得到的数值型编码数据。Excel 文件是标准的数据表文件（spreadsheet datafiles）。DL 文件包含的数据与 Raw 文件一样，但是额外包含关于数据的一些信息，如行数、列数、变量名、标签等。如果从头开始键入数据，建议利用 DL 文件类型，因为这种文件易于转化或者输出成为其他类型的文件。

一、初始数据文件

初始数据文件（raw data file）仅仅包含数字，不包含数据的行数、列数、标签、标题等信息，因而只能以矩阵的形式键入。在 UCINET 6.504 中，点击 File→Text Editor 或者直接点击主界面的第五个快捷图标“Edit Text File”，即可打开文本编辑器，可以在其中直接输入初始数据。假设输入四个点之间的关系数据：

```
0 1 1 0
1 0 1 1
1 1 0 0
0 1 0 0
```

输入完毕之后保存为纯文本文件，并命名为 4nodes。然后，在 UCINET 中，点击 Data→Import Text File→Raw Matrix，在出现的对话框中，选择 4nodes 并点击 OK 后，即可将该文本文件中的数据转换为 UCINET 数据。尽管这种程序比较

方便,但是我们不建议使用它。原因之一在于,计算机不能对数据进行检查:如果第一行恰恰缺少了一个值的话,程序本身是不知道的。原因之二在于,数值没有标签,无法用来确定各个点到底是谁。

二、Excel 文件数据

如果初始的网络数据是用 Excel 格式保存的,UCINET 可以打开 Excel 文件,这需要利用 Data→Import Excel→Matrices 命令。需要注意的是,Excel 数据文件虽然可以有上万行,但是数据的列数最多只能支持 255 列。因而,不能用 Excel 来存储超过 255 列的大型网络数据,换句话说,用 UCINET 打开的 Excel 文件最多有 255 列。

三、数据语言类型的文件

在 UCINET 软件中,常见的数据输入方法是在一个文本文件中输入数据,这可以利用任何一种字处理程序,如写字板或者 Microsoft Word。如果该文件要用 dl 开头,则表明它是数据语言(Data Language)文件(DL file),它包含一系列数据、描述数据的关键词和语句等,这些都是关于数据的基本信息。用 UCINET 本身的 File→Text Editor 编辑器可以输入多种类型的数据语言文件,其中的关键词"dl"(代表 data language)是必需的,它界定了该文件是一个数据语言文件,并且必须放在文件的最前端。输入后保存为纯文本文件,然后利用 Data→Import Text File→DL 即可将文件转换成为 UCINET 本身的数据。下面介绍多种可以输入的数据语言类型文件。

(一) 全矩阵格式的方阵数据

在 UCINET 中,打开 File→Text Editor 编辑器后,可以输入全矩阵形式的 1-模数据文件。就上述四个行动者之间关系的简单数据来说,其对应的数据语言文件可在 Text Editor 编辑器中,按照如下方式输入全矩阵格式(full matrix format):

```
dl n = 4 format = fullmatrix
data:
0  1  1  0
1  0  1  1
1  1  0  0
0  1  0  0
```

关键词"dl"必须放在文件的最前端。语句"n＝4"表示该矩阵拥有 4 行和 4 列，即这是一个方阵数据。等号可用一个或者多个空格或者逗号代替。因此，"n＝4"，"n 4"，"n，4"这些表达方法都是等价的。语句"format＝fullmatrix"表明数据以一种常规矩阵的形式输入(与下文将介绍的 linkedlist，lowerhalf matrix 等形式不同)。由于全矩阵是默认选项，因此，该语句是需要加以选择的。关键词"data："表示标题信息(关于数据的信息)结束，下面就是数据本身了。因此，关键词的顺序很重要。必须记住"dl"词是第一个，随后的信息都与矩阵的维度有关，再接下来的信息是其他关键语句，随后才是"data："这个关键词。

在输入数据的时候要注意标点符号。一般来讲，冒号(full colon)"："表示下面紧接着有一些信息，如一系列数据取值或者标签。分号(semicolon)"；"或者回车键(重启一行)意味着一条语句的结束。每个数据都必须用一个或多个空格或回车键相互隔离开。全部非数值型数据都被认为是缺失值。数据的格式未必要求行数和列数相等，只要全部值都出现，从左到右的顺序，从上到下的排列即可。下面的输入方法就是可以接受的：

```
dl n = 4 data:
0 1 1 0 1 0 1 1
1 1 0 0 0
1 0 0
```

在全矩阵形式中，我们必须将邻接矩阵全部键入。矩阵的各项值之间必须至少空一格，一定不要用逗号或者其他符号。

(二) 长方形矩阵

长方形矩阵(rectangular matrices)可按照如下方法输入，即指定矩阵的行数和列数。如下所示：

```
dl nr = 6, nc = 4
data:
0 1 1 0
1 0 1 1
1 1 0 0
0 1 0 0
1 0 1 1
1 1 0 0
```

“nr=6”表示矩阵包含 6 行(其中,nr 等于 number of rows 的缩写),“nc=4”表示矩阵包含 4 列(nc 是 number of columns 的缩写)。当然,如果你的数据是 18 行 7 列的话,则应该键入“nr=18, nc=7”。

(三) 加入标签

DL 文件也可以包含行动者的标签(labels)。可以在标题行中加入标签信息,标签的长度最长为 18 个字符,不能包含空格或者逗号(注意,标签如用中文,UCINET 在打开文件的时候可能出现乱码。因此,笔者建议用英文符号表示标签)。如下所示:

```
dl n = 5
format = fullmatrix
labels:
Andy, barry, chris, david, emma
data:
0  1  1  1  0
1  0  0  0  1
1  0  0  1  0
1  0  1  0  1
0  1  0  1  0
```

“labels:”这个关键词表示下面给出行和列的标签。一个标签中间不能包含空格,除非用引号括起来,如“Tom Smith”。另外要注意的是,“labels”这个词后面应该跟着一个冒号。

也可以在不同行上输入标签,如下所示:

```
dl n = 5
format = fullmatrix
labels:
barry, david, lin, pat
russ
data:
0  1  1  1  0
1  0  0  0  1
```

```
1  0  0  1  0
1  0  1  0  1
0  1  0  1  0
```

标签还可以嵌入到数据之中，这需要用到 embedded 命令，如下所示：

```
dl n = 5
format = fullmatrix
labels embedded
data:
larry david lin pat russ
Larry 0  1  1  1  0
david 1  0  0  0  1
Lin 1  0  0  1  0
Pat 1  0  1  0  1
Russ 0  1  0  1  0
```

在输入上述数据之后，需要在研究者指定的盘面上保存该文件成为纯文本文件，并且给该文件命名。如果将该文件中的信息转换为 UCINET 数据，需要在 UCINET 中，沿着 Data→Import Text Files→DL 路径，选择刚刚保存的文本文件即可打开成为 UCINET 形式的数据文件。

也可以针对行和列分别指定标签，当矩阵不是方阵的时候更需要如此。例如，

```
dl nr = 6, nc = 4
col labels:
hook, canyon, silence, rosencrantz
data:
0  1  1  0
1  0  1  1
1  1  0  0
0  1  0  0
1  0  1  1
1  1  0  0
```

0

另外一种输入标签的方法是将标签作为输入的数据矩阵的一部分,如下所示:

```
dl nr = 6, nc = 4
row labels embedded
col labels embedded
data:
        Dian Norm Coach  Sam
Mon      0    1    1     0
Tue      1    0    1     1
Wed      1    1    0     0
Thu      0    1    0     0
Fri      1    0    1     1
Sat      1    1    0     0
```

"row labels embedded"和"column labels embedded"这两条语句表明行标签和列标签嵌入数据本身。也可以简单地写成"labels embedded",以便表明行标签和列标签都可以在数据中找到。

(四) 多个矩阵的输入

有时候要用一个文件来表征多个矩阵(multiple matrices)。例如,我们可能测量一群人之间的多种关系。具体做法如下:

```
dl n = 4, nm = 2
labels:
GroupA, GroupB, GroupC, GroupD
matrix labels:
Marriage, Business
data:
0  1  0  1
1  0  0  0
0  0  1  0
1  0  0  1

0  1  1  1
```

```
1 0 0 0
1 0 0 1
1 0 1 0
```

“nm=2”表示文件包含两个矩阵(number of matrices=2)。关键词“matrix labels:”表示下面的两个词(“marriage”和“business”)是两个矩阵的名称。

(五) 对角线缺失的情况

在默认情况下,程序假设数据是“全矩阵形式”(fullmatrix format),其中不存在缺失值。然而,在方阵情形中,有时候忽略某些数据会带来方便。例如可以忽略对角线的值(diagonal absent)。此时数据的格式为:

```
dl n = 4
diagonal = absent
labels:
Sanders, Skvoretz
S.Smith, T.Smith
data:
  1 1 0
1   1 1
1 1   0
0 1 0
```

利用 Data→Import Text File→DL 程序打开该文本时自动地将对角线的值设定为缺失值。

(六) 半矩阵数据的输入

上面的矩阵是对称矩阵。而在输入对称数据的时候,不必把全部数据都输入,只需要输入方阵的下半部分或上半部分(lowerhalf and upperhalf matrices)即可。例如,上例中的数据也可以按照下面的方式输入,即仅输入矩阵的“下半部分”。

```
dl n = 4
format = lowerhalf
diagonal = absent
```

```
labels:
Sanders, Skvoretz
S.Smith, T.Smith
data:
1
1  1
0  1  0
```

UCINET 会自动加入矩阵上半部分的值,对角线的值也是缺失值。把"lowerhalf"用"upperhalf"代替,就可以输入矩阵上半部分的值。注意,如果不加入"diagonal absent"这个语句,那么应该指定对角线的值。例如,

```
dl n = 4
format = lowerhalf
labels: Sanders, Skvoretz, S.Smith, T.Smith
data:
2
1  2
1  1  2
0  1  0  2
```

(七)块矩阵格式

另外一类数据格式可用来创建目标矩阵,即"块矩阵"(blockmatrix)。在这种形式中,用不着直接输入数据,输入创建数据的方向即可。例如,为了输入如下所示的高度结构化的矩阵:

```
0  1  1  1  1  0  0  0  0  0
1  0  1  1  1  0  0  0  0  0
1  1  0  1  1  0  0  0  0  0
1  1  1  0  1  0  0  0  0  0
1  1  1  1  0  0  0  0  0  0
0  0  0  0  0  0  1  1  1  1
0  0  0  0  0  1  0  1  1  1
```

0 0 0 0 0 1 1 0 1 1
0 0 0 0 0 1 1 1 0 1
0 0 0 0 0 1 1 1 1 0

我们可以创建如下所示的文件：

```
dl n = 10 format = blockmatrix
data:
rows 1 to 10
cols 1 to 10
value = 0
rows 1 to 5
cols 1 to 5
value = 1
rows 6 7 8 9 10
cols6 to 10
value = 1
diagonal 0
value = 0
```

在 blockmatrix 格式中，需要首先界定一个集合的元素，然后赋予一个值。在例子中，前三行把矩阵所有值都赋予 0，接下来的三行把矩阵左上半部分的值赋予 1，随后的三行又把矩阵右下半部分的值赋予 1，最后两行给予主对角线或者 0 对角线上每个值赋予 0。如果第三个主对角线上的值都是 5，则需要将最后两条命令修改为：

```
diagonal3
value = 5
```

如果再打开文件，会出现如下形式：

0 1 1 5 1 0 0 0 0 0
1 0 1 1 5 0 0 0 0 0
1 1 0 1 1 5 0 0 0 0
1 1 1 0 1 0 5 0 0 0

1	1	1	1	0	0	0	5	0	0
0	0	0	0	0	0	1	1	5	1
0	0	0	0	0	1	0	1	1	5
0	0	0	0	0	1	1	0	1	1
0	0	0	0	0	1	1	1	0	1
0	0	0	0	0	1	1	1	1	0

再举一例。请看如下矩阵：

100	0	0	0	0	0
90	100	0	0	0	0
80	90	100	0	0	0
70	80	90	100	0	0
60	70	80	90	100	0
50	60	70	80	90	100

输入上述矩阵,需要键入如下命令行：

```
dl n = 10 format = blockmatrix
data:
rows all
cols all
value 0
diag 0
val = 100
diag  - 1
val = 90
diag  - 2
val = 80
d  - 3
v = 70
d  - 4
v = 60
```

```
d -5
v = 50
```

需要注意的是,关键词“all”表明全部的行和列。还要注意的是,行(rows)、列(cols)、取值(value)和对角线(diagonal)都可简写为前几个字母。例如,“d-5”的含义是“在主对角线以下第五个对角线”,“v=50”表示“令这些元素的值都等于 50”。

(八) 关联列举格式

顾名思义,在关联列表形式(linked list formats)的数据输入方法中,使用者只需要指定数据中实际有关联的关系,忽略无关的关系。这些数据输入的形式在数据较大并且联系比较松散的时候是很方便的。

关联列表形式又分为两类:点列表(nodelists)和边列表(edgelists)。它们又各分为两小类,一小类对应于 1-模数据,另一小类对应于 2-模数据。只有边列表格式(edgelists)允许使用多值数据。

1. 点列表形式(Nodelist Format)。

点列表是由连接一个既定点的一系列点构成的。它又分为两小类:nodelist1 和 nodelist2。前者用于输入 1-模方阵网络数据,后者用于输入长方形的 2-模矩阵数据,如一个案例—变量矩阵(a case-by-variable matrix)或者一个“项目—使用矩阵”(an item-by-use matrix),这两种形式都能够接受字母数据。

(1) 点列表格式-1(nodelist1)。这种格式只能用来输入 1-0 矩阵数据。在点列表格式-1 中,可以给出一个点的名字或者数目(name or number),随后给出与该点相连的其他点的名字或数量,如下所示:

```
dl n = 4, format = nodelist1
labels:
Sanders, Skvoretz, S.Smith, T.Smith
data:
1  2  3
2  1  3  4
3  1  2
4  2
```

“format＝nodelist1”这个语句指定一种形式,即数据中每行的第一个数字确定了一个行动者(称之为自我点),其关系指向该行中其他点。例如,第三行数据指出,行动者 3(S.Smith)的关系指向 Sanders 和 Skvoretz。各个“自我点”的顺序(即各行的顺序)不重要,一行中的各个他者的顺序也不重要。不包含他者的自我点可略去(但是在“n＝”语句中可不能略掉),但是要在标签中指出来。

如果一行中有太多的他者需要输入,可以为该行动者创建第二行,但是新一行中的第一个数字必须说明这些他者的关系指向谁。例如,下面的语句等同于上述语句:

```
dl n = 4, format = nodelist1
labels:
Sanders, Skvoretz, S.Smith, T.Smith
data:
1  2  3
2  1  3
2  2  4
3  1  2
4  2
```

也可以直接输入标签,从而产生与上述同样性质的数据:

```
dl n = 4, format = nodelist1
labels embedded
data:
sanders   skvoretz s.smith
skvoretz  sanders s.smith t.smith
s.smith   sanders skvoretz
t.smith   skvoretz
```

可以既在数据的左半部分列举出所有的标签(在“labels:”语句之下),又让标签嵌入数据中(前提是保证“labels embedded”命令存在)。这种过程的好处在于,作为结果的矩阵的各行和各列将以标签指定的顺序出现。例如,下面的语句

将产生一个其行和列按照字母顺序排列的矩阵。

```
Dl n = 4
format = nodelist1
labels:
Alfred, Betty, Chuck, David
Labels embedded
Data:
Betty chuck alfred
Alfred chuck david
David chuck betty
Chuck betty
```

如果不列举出标签，那么在结果矩阵中将根据指定的标签顺序排列：第一行/列将是 Betty，第二行/列将是 Chuck，第三行/列是 Alfred，依次类推。

依次点击 File→Text Editor 编辑器，键入如下命令，即可输入上述“4nodes”数据。这个文件（纯文本文件）表示下面的矩阵（format＝nodelist1 的含义即表明这是一种列举点格式的输入形式）：

```
DL n = 4
format = nodelist1
data:
1  1  2  3
2  1  2  4
4  1  4
```

输入数据后，保存为纯文本文件（假设该矩阵保存在 D:\datafiles\文件夹中）并用“4nodes”命名。然后，点击 Data→Import Text File→DL，找到 D:\datafiles\文件夹中的“4nodes”纯文本文件，点击 OK 后，即得到了 UCINET 数据矩阵如下。

```
Input file:           D:\datafiles\4nodes.txt
Output datatype:      Real
Output dataset:       4nodes(D:\datafiles\4nodes)
```

```
     1     2     3     4
   ----  ----  ----  ----
1  1.00  1.00  1.00  0.00
2  1.00  1.00  0.00  1.00
3  0.00  0.00  0.00  0.00
4  1.00  0.00  0.00  1.00
```

(2) 点列表格式-2(nodelist2)。nodelist2 格式用于输入 2-模数据。如下例所示:

```
dl nr = 3, nc = 4 format = nodelist2
row labels embedded
column labels embedded
data:
benny sex_roles, stats
ginny aging, intro
sally stats, intro, aging
```

该例子记录了三个人教四门课的情况。数据第一行表明,Benny 教性别角色(sex roles)和统计学(stats)这两门课程。第三行说明,Sally 除了教授统计学之外,还讲解导论(intro)和老年社会学(aging)这两门课程。每一行的第一个选项总是代表一个行项(row entity),该行中其他项目代表列项(column entities)。

(3) 等级列表格式(Rankedlist Formats)。等级列表格式是对点列表格式的简单变体,它假定与某一特定点有关系的其他点要以一种有意义的方式输入。该程序然后根据这种顺序赋予关系强度。适用于输入 1-模网络数据的等级列表格式为等级列表格式-1(rankedlist1)。例如,请看如下文件:

```
dl n = 4, format = rankedlist1
labels embedded
data:
sanders   skvoretz s.smith
skvoretz   sanders s.smith t.smith
s.smith   sanders skvoretz
t.smith   skvoretz
```

如果我们看数据语句(data:)下面的第一行(即 sanders skvoretz s.smith),它说明,行动者 Sanders 将 Skvoretz 评为一级(编码为 1),将 S.Smith 评为二级(编码为 2)。根据默认值,没有被列举出来的行动者被赋予 0,除非在 dl 语言中语句"recodena=no"出现在数据陈述一项的上面。读取该文件将得到如下矩阵:

	Sanders	Skvoretz	S.Smith	T.Smith
Sanders	0	1	2	0
Skvoretz	1	0	2	3
S.Smith	1	2	0	0
T.Smith	0	1	0	0

与点列表情况完全一样,在输入 2-模数据的时候也要用到等级列表格式-2(rankedlist2)。其格式如下:

```
dl nr = 3, nc = 4 format = nodelist2
row labels embedded
column labels embedded
data:
benny sex_roles, stats
ginny aging, intro
sally stats, intro, aging
```

在这种情况下,"benny sex_roles, stats"对于矩阵中 benny 所在行来说,赋予"sex_roles"的值为 1,赋予"stats"的值为 2。

2. 边列举格式(Edgelist Formats)。

如果数据本身比较疏松,联系比较松散,那么可以仅仅输入有关联的行动者对。边列举格式并不列举出每个自我点的一系列他者,而是具体指定每一种关系(或者矩阵的每个元素)。它又分为两种:edgelist1 和 edgelist2。前者应用于 1-模数据矩阵;后者用于 2-模数据矩阵,如案例—变量矩阵(case-by-variable matrix)。这两种格式都可输入字母型数据。

(1) 边列表-1 格式。在边列表-1 形式(edgelist1 format)中,要有一系列边(edges)及其取值。每行数据都是一个有序行动者对(ordered pair of actors),可选择在其后标记一个值,以便表明关系的强度。如下所示。

```
DL n = 5
format = edgelist1
labels:
george, sally, jim, billy, jane
data:
1  2
1  3
2  3
3  1
4  3
```

标签也可以直接加入到数据本身之中，如下所示：

```
DL n = 5
format = edgelist1
labels embedded:
data:
george sally
george jim
sally jim
jim george
billy jim
```

各个点根据前后顺序进行排列。为了控制在结果数据矩阵中点的顺序，也可以仅在标题行中加入标签，或者加入到数据之中。如下所示：

```
DL n = 5
format = edgelist1
labels:
george, sally, jim, billy, jane
labels embedded:
data:
george sally
```

```
george jim
sally jim
jim george
billy jim
dl n = 4 format = edgelist1
labels:
Sanders, Skvoretz, S.Smith, T.Smith
data:
1  2  1
1  3  2
2  1  1
2  3  1
2  4
3  1  1
3  2  na
4  2  10.3
```

在该例子中,数据的第四行告诉我们,行动者 2 以强度 1 与行动者 3 关联在一起。如果关系强度被忽略,如第五行所示,那么将假定它为 1。如果忽略了一对行动者(dyad),如(4,3)这一对所示,就假定二者之间的关系强度为 0。关系强度可以是整数或者实数。非数值型的值(non-numeric values)如"na"被编码为缺失值(missing)。从这种输入数据中生成的邻接矩阵如下所示:

	Sanders	Skvoretz	S.Smith	T.Smith
Sanders	0	1	2	0
Skvoretz	1	0	1	1
S.Smith	1		0	0
T.Smith	0	10.3	0	0

边列表-1 格式的一个特点是可以处理字母数据(character data)。因此允许"labels embedded"这样的语句出现,如:

```
dl n = 3 format = edgelist1
      labels embedded
      data:
      sanders skvoretz 1
      sanders s_smith 1.782
      skvoretz sanders 1.09
```

这种句法将从每个数据记录的前两项中提取行和列的标签，并按照它们出现的顺序进行排列。因此，在本例子中，矩阵拥有 3 行 3 列，第一行/列是 Sanders，第二行/列是 Skvoretz，第三行/列是 S_Smith。该程序将处理任意长度的字母标示(character identifiers)，但是只有前 19 个字母(或者 255 个，如果 Options→longlabels 得到激活)才能被存储起来作为行或列的标签。

如果加入"alpha=no"这样的语句，行和列将根据输入数据的顺序排列。在上述例子中，第一行是 sanders，第二行是 skvoretz，第三行是 s_smith。

在一个文件中也可能拥有多种关系，这只需要用一个竖杠(|)表示即可。例如，

```
dl nm = 2 n = 3 format = edgelist1
labels embedded
data:
sanders skvoretz 1
sanders s_smith 1.782
skvoretz sanders 1.09
sanders s_smith 1
skvoretz sanders 1
```

(2) 边列表格式-2。边列表格式-2(edgelist2)应用于输入 2-模网络数据。例如，考虑下面的矩阵：

	COHESION	SIMILARITY	CENTRALITY
CLIQUE	1	0	1
CLOSENESS	1	0	1
REGE	0	1	0

DENSITY	1	0	1
CONCOR	1	1	0

用边列表格式-2 编辑的输入文件如下所示：

```
dl nr = 5 nc = 3, format = edgelist2,
labels embedded
data:
clique cohesion
clique centrality
closeness cohesion
closeness centrality
rege similarity
density cohesion
density centrality
concor cohesion
concor similarity
```

在该例子中，全部关系线的权重都是 1。而下面的关系数据则有不同的权重。

```
dl nr = 2, nc = 3 format = edgelist2
labels embedded
data:
sanders skvoretz 1
sanders s_smith 1.782
skvoretz sanders 1.09
```

尽管这种数据与边列表-1 格式的数据一样，输出的结果却不同：

	Skvoretz	S_Smith	Sanders
Sanders	1	1.782	0
Skvoretz	0	0	1.09

需要注意，只有两个不同的对象(Sanders 和 Skvoretz)出现在发生阵的行中(看每一数据纪录的第一项)，而三个不同的对象(Skvoretz, S_Smith 和 Sanders)出

现在发生阵的列中(看每一数据纪录的第二项)。还要注意到如下事实,即在该例子中行对象的名称恰恰与某些列对象的名称一样,而某些列对象没有任何含义,因此该程序不认为它们是一样的。

(3) 边阵列格式 1(The EdgeArray1 Format)。这种格式与边列表格式很相似,但是它允许在各种关系上拥有多种属性(即多种关系)。例如,

```
dl nm = 4 n = 3 format = edgearray1
labels embedded
matrix labels:
avgcalls
yearstogether
pctincommon
kin
data:
sanders skvoretz  1      62  33.7  0
sanders s_smith   1.782  29  66.2  1
skvoretz sanders  1.09   59  99.0  0
```

按照这种方式输入数据,得到的结果为四个矩阵。

四、UCINET 数据表编辑器

在 UCINET 中,我们可以直接键入数据,也可以剪切、粘贴数据到 UCINET 数据表中。只要点击 UCINET 主界面中的数据编辑器快捷方式(matrix spreadsheet)或者点击 Data→Spreadsheet 就可以打开 UCINET 的编辑器。如果你希望直接键入数据,那么可以在编辑器中键入。你也可以利用"复制"、"粘贴"等方式,将 Excel 中的数据拷贝到 UCINET 中;或者利用 Ctrl+C,再选择 UCINET 中的 Edit→Paste 命令。在 UCINET 数据表中输入数据后,需要保存为 UCINET 文件。需要注意的是,如果你已知数据的规模,那么首先要加以指定。如果数据是对称的,那么在用数据表(spreadsheet)输入任何数据之前点击"对称"(Symmetric)模式按钮,这样就会自动在数据的另一半填上数据,你只需要在数据表中输入非零值即可。一旦输入结束,可以点击 Fill 按钮,那么所有的元素就被赋予 0 值。如果你的数据拥有多种关系,那么利用数据表本身的工具

栏中 Edit→Insert Sheet 加入额外的矩阵(利用 Edit→Delete Sheet 可删除关系矩阵)。网络中的每个矩阵都可用“重命名”来命名,这要用到 Edit→Rename Sheet 命令。

第三节　UCINET 数据的预处理

整体网数据的处理方法取决于要研究的具体问题。换句话说,整体网数据的处理方法因问题不同而不同。不过,在正式处理数据之前,有时候需要进行数据的预处理,从而满足进一步分析的需要。需要数据预处理和转换的情况较多。一种情况可能在于,数据收集过程中存在缺陷,需要弥补。例如,如果研究一个群体中每个人之间相互评价的“亲密关系”,其结果可能是,某些较低的评价值(如小于 10)本质上是随机的。在这种情况下,你就有必要把所有低于 10 的值进行重新编码,都变成一个值,如 0。

另外一种情况在于,不同的程序要求不同的数据处理(data processing)和转换计算。例如,如果收集到多人之间的多值朋友关系数据,并且希望界定甲和乙所在的派系,则应该对数据进行二值化处理,因为派系程序只能针对二值数据进行分析。

数据预处理的第三种情况在于创建新变量,从而更便于分析。例如,假设你拥有一些行动者之间的态度数据,也有他们之间的关系强度数据。一个自然的假设就是检验态度上的相似与关系的强度是否相关,可以利用 QAP 程序来检验这个假设。然而,应该首先获得行动者之间在态度上的相似性,这可以通过计算行动者—态度矩阵(actors-by-attitudes matrix)中各行之间的相关性来获得数据。

UCINET 6 提供了大量数据转换和预处理程序。它们生成新的数据,可以进行新的分析。本节介绍如何对初始数据进行预处理。

一、抽取子图和子矩阵

把一个点或多个点从一个网络中移出去,这可能是最常见的数据调整过程,这个过程对应于从一个邻接矩阵中转移出一行(列)或者多行(列)。在 UCINET 6 中最简单的做法就是运行 Data→Extract 程序,它会从一个输入矩阵中取出一个矩阵。作为输入的文件当然应该是多元的矩阵文件,需要指定保

留(删除)哪些行和(或)列。如果仅仅去掉一行(列)或两行(列),就应该指出这些行或列。在执行 Data→Extract→Extract Submatrix 程序的时候,需要你自己指定如下信息:

```
| Input dataset:                              |
| KEEP or DELETE?          KEEP               |
| Which rows               ALL                |
| Which columns            ALL                |
| Which matrices           ALL                |
| Output dataset:          EXTRACT            |
```

“Which rows?”要求具体指定哪些行的行动者,诸如“1 To 10, 12 To 24”等。对于列也必须进行同样的操作,在此案例中只需要键入“same”即可。如果数据有标签,并且你知道行动者的标签而不知道其代码,那么点击所出现的对话框中的“L”按钮,就可以从列表中选择适当的行动者。在按住 Ctrl 的同时,依次点击需要的点,就可以激活多个行动者。你也可以为一个包含群体指标矩阵(group indicator matrix)的 UCINET 6 数据命名,该数据可能是根据一些分析程序(如 Cliques)输出的。所谓群体指标矩阵(A group indicator matrix),指的是一个二值的群体—行动者矩阵(binary group-by-actor matrix),该矩阵中元素之值的含义是:如果行动者 j 是群体 i 的一个成员,则 X(i, j)=1,否则 X(i, j)=0。例如,本章第一节介绍的派系分析(Network→Subgroups→Cliques)程序会把生成的群体指标矩阵命名为 CliquesSets;如果你希望在 CliquesSets 中取出一个只由第六个派系的点构成的子图,需要在“Which rows?”后面输入 CliquesSets row 6。然后,针对对话框中列的处理也类似,或者利用“same”,或者键入“CliqueSets row 6”。该语句的含义是,“我所需要保留的行动者是在一个名为 CliquesSets 的 UCINET 6 数据中第六行的行动者”。特别要注意的是,如果你的数据文件名中包含空格,如“My Clique Sets”,则需要在文件名上加括号,如下所示:

Which rows to KEEP? “My Cliques Sets” row 6

二、数据的合并

有时候需要在不同的文件中保存同一群行动者之间的多类关系数据,有时

候又需要把它们合并(merging)在一个文件之中。例如,我们可能针对单独一类数据进行分析,但是对于某些程序(如位置分析 positional analysis)来说,可以针对有关该数据的全部关系矩阵(例如,友谊关系、借款关系、亲属关系)同时进行分析,这就需要将数据合并在一起,Data→Join 程序就可以完成数据合并的任务。合并的角度有三个:行合并、列合并、层次合并(即矩阵合并)。例如,请看如下两个矩阵:

矩阵 1

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

矩阵 2

2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

如果根据"行"合并,就是行数相加,即把第二个矩阵加在第一个矩阵的底部,从而产生一个 8 行 4 列的新矩阵。如何根据"列"合并,就把第二个矩阵加在第一个矩阵的右侧,从而产生一个 4 行 8 列的新矩阵。如何根据"层次"(levels)合并,即将第二个矩阵和第一个矩阵放在一起,从而在一个数据中产生两个 4 行 4 列的新矩阵。这三种合并的结果分别如下所示。

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

行合并的结果

1	1	1	1	2	2	2	2
1	1	1	1	2	2	2	2
1	1	1	1	2	2	2	2
1	1	1	1	2	2	2	2

列合并的结果

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

层次合并的结果

三、置换与排序

对矩阵的行和列重新排序,可以使得同一个群体的成员相互邻接,进而容易把握矩阵的整体结构。UCINET 6 中区分出两类排序方式:置换与排序(permutations and sorts)。

重排是使用者具体指定点的新顺序的方法。例如,可令 Permute 程序指定点的顺序为"1, 6, 2 to 5, 10 9 8 7"。例如,下面就是 Data→Permute 程序的输入格式:

Input dataset:	
New order of rows:	
New order of cols:	
New order of matrices:	
Output dataset:	Permuted

重排方法对指定的行和列的顺序同时进行调换。相比之下,分类(sorts)方法则对全部点的顺序进行重新安排。例如,我们可能要求程序根据中心度的降序排列来对各个点进行重新排序。要做到这一步,我们必须首先拥有一个关于各个点的中心度的数据。然后,告诉程序哪个数据包含中心度向量,要具体告诉程序该向量存在于哪个数据的哪一行或者哪一列。

四、转置和改型

有时候需要把矩阵的行和列进行调换,这涉及矩阵的转置和改型(transposing and reshaping)。在有关世界体系研究的文献中,一般约定网络数据中的值 X(i, j)代表从国家 j 向国家 i 输出的货物值。对于网络分析者来说,这可能令人迷惑,因为在网络文献中,X(i, j)总表示从 i 指向 j 的值。在这种情况下,需要做的工作只能是对矩阵进行转置处理,要利用 Data→Transpose 程序来完成。

如果你有一个三维数据(即在一个单一数据库中的多个矩阵),Data→Transpose 程序会对每个矩阵进行转置处理。然而,如果你不希望对行和列进行转置,而希望对列和层次(level)进行转置的话,又该怎样做呢?例如,在 UCINET 中有一个多元网络数据 Newcomb 是由 12 个 17×17 矩阵构成的,每一个矩阵针对的都是一周内收集到的信息。行代表个体(回答者),列代表他者,层次(levels)代表各个周。数据中的取值代表每个个体在每一周对每个他者的评价等级。假设我们需要对数据重新组织,从而使得某个回答者对他者在各周的评价等级是一个挨着一个的(因此我们可以看出这些值是升高还是降低)。这样的话,就需要用矩阵的列代表周,行代表个体,各个层次指代他者。因此,下一步需要对初始矩阵的列和层次互换(转置)。这一点可通过矩阵代数算法(Tools→

Matrix Algebra)完成,如下所示:

tnewc1=transp(newcomb1 columns levels)

该语句的含义是,“对数据库 Newcomb1 中的列和层次进行互换,将结果保存在 tnewc1 中”。

五、重新编码

有时候需要重新编码(recode)。例如,有时候可能把从 0 到 9 的全部值都变为 0,从 10 到 19 的全部值都变为 1。这种想法可利用 Transform→Recode 程序来实现。一旦找出有待重新编码的输入文件之后,即可进行如下工作:

Values	0	to	9	are recoded as	0
Values	10	to	19	are recoded as	1
Values	20	to	100	are recoded as	2
Values	na	to	na	are recoded as	0
Values		to		are recoded as	

要想输入上述类型的数据,可利用 tab 键或者鼠标从左到右、从上到下移动。要注意的是,在第四行中,命令要求缺失值(missing values)(用 na 表示)变为 0。

再编码(recode)命令比较有用,可以把 0 转换成缺失值,反之亦然。然而需要注意的是,你不能引导程序只改变主对角线上的值。要想这样做,可利用 Transform→Diagonal 程序,该程序允许改变主对角线上的值。

有时候 UCINET 程序只针对二值数据进行分析,这就需要把矩阵中全部数值二值化为 0 或 1。此时程序将自动地把大于 0(或者研究者指定)的值变为 1。Transform→Dichotomize 程序将把每个元素与你提供的切割值进行对比,然后进行二值化处理。

另外一类重新编码操作的就是取相反数(reversal of values)(Transform→Reverse)。例如,对于一个取值为 0 和 1 的二值矩阵来说,这种操作意味着把所有的 1 变为 0,把全部 0 变为 1。如果关系是从“喜欢”转向“讨厌”的话,那么这种转换是有用的。对于多值数据来说,该程序将用最大值减去每个值。例如,对于数据取值为 0, 1 和 2 来说,该程序计算的结果为将 0 改为 2,将 2 变为 0,将 1 也变为 1。

六、线性转换

实际上,所谓的转换运算(reversal)就是一种线性变换(linear transformation) Y(i, j)=X(i, j)+Min(X)−Max(X)。为了执行更一般的转换(如 Y(i, j)=bX(i, j)+m,其中 b 和 m 是常数项),就有必要利用矩阵代数程序了。例如,为了把一个矩阵 DAVIS 中的全部值都乘以 7,应该点击 Tools→Matrix Akgebra,在出现的对话框中键入:

davis7=linear(davis 7 0)

这意味着,"把矩阵 DAVIS 中的每个值都乘以 7,然后加上 0,计算结果命名为 davis7"。为了把矩阵中的全部值除以 2,并减去 1,需要键入:

davis7=linear(davis 0.5−1)

七、对称化处理

把一个矩阵进行对称化处理(symmetrizing),就是对于全部 *i* 和 *j* 来说,要求 X(i, j)等于 X(j, i)。在很多情况下确实需要对称化处理。一种情况是关系本身具有对称性,如"与谁共进午餐"关系就是这样,但是由于测量的误差,实际的数据可能不对称。另外一种情况是,一些非对称关系(如借钱给某人)可能扩展为"金钱关系",即变成对称关系。在这些情况下,就需要令 X(i, j)等于 X(j, i)。

UCINET 提供多种对称化处理的程序,包括用均值、最小值或者最大值代替矩阵元素之值。还有一类不常见的方法是,如果 X(i, j)=X(j, i),则设定 X(i, j)和 X(j, i)都等于 1,否则都设定为 0。实际上,这种算法会生成一个新的矩阵,表明每一对点是否一致。矩阵中所有值的均值将给出一致性的二人组所占比例。对称化处理需要用到 Transform→Symmetrize 命令。

八、捷径距离和可达性

捷径距离矩阵(geodesic distance matrix)表述的是各个点之间的捷径距离。捷径距离矩阵可被看成是一种影响(凝聚)指数。

为了计算一个捷径距离矩阵,需要运行 Network→Cohesion→Geodesics 步

骤。在 UCINET 6 中，捷径距离的概念得到了推广，可以对各种类型的多值矩阵进行分析。例如，你的数据可能记录的是朋友关系的强度。从点 a 到 b 到 c 到 d 的途径的强度界定为它们之间的最弱关系链（weakest link between them）。从任何一点到其他点的捷径路径（geodesic path）定义为二者之间的最强路径。两点之间的捷径距离定义为二者之间的强路径中的连线数。另一类多值数据是由“费用”（costs）或者与一对点相关联的距离构成的。在这种情况下，捷径距离定义为点与点之间的最短途径距离，要根据途径中的每条线值之和来计算。

与“距离”密切相关者是“可达性”（reachability）。在一个邻接矩阵 X 中，该概念常常指的是一种数据的传递性闭圈（transitive closure），也就是说，如果从 i 到 j 之间存在任何一个途径的话，那么 Y(i, j)＝1。然而，UCINET 6 推广了该概念，使之适应于各种类型的多值矩阵。例如，对于一个记载着强度的矩阵来说，i 和 j 之间关系的能力或频次（capacities or frequencies）就是 i 到达 j 的最强途径的强度。如果关系强度值是二值化的 1 和 0，那么任何两点之间的关系强度的最大值就总是 1，除非二者之间根本不存在途径。这就产生了“可达性”（reachability）概念。对于成本数据（cost data）来说，可达性被界定为最便宜途径的成本（the cost of the cheapest path）。对于概率数据来说，可达性被界定为两点之间最可能出现的途径的概率。因此，可达性矩阵可用作一种测度，来测量凝聚力或者期望影响力（expected influence）。

九、矩阵的组合

利用组合（aggregation）命令可以把同一群行动者之间的多种关系组合在一起。例如，对两个行动者之间的“影响”、“互依性”和“关系强度”等三种关系上进行测量得到的关系总数可能是 0 到 6。一些学者称之为关系的“多丛性”（multiplexity），但是 UCINET 称之为关系的组合（combination of ties that bind two actors）。

问题是，如何测量两个行动者在多种关系上的属性？在一个数据“内部”，有两种方式可用来加总关系。第一种方式是利用前文介绍的 data→CSS。CSS 程序中又有多种加总的算法，如选择“分片”（slice）法。第二种方法是矩阵代数算法（在 UCINET 中，选择 Tools→Matrix Algebra）。例如，在 UCINET 自带的数据库中有一个数据为 PADGETT，其中包含两种关系。为了对它进行汇总，

需要打开 PADGETT 数据，在 UCINET 中矩阵算法(Tools→Matrix Algebra)的输入命令(Enter command here)栏键入：

agpadge=total(padgett rows cols)

该命令的含义是，对数据 padgett 的各个值进行汇总(sum the values of dataset)，并将结果保存为数据库 agpadge。如果我们还希望针对各个列进行加总，则可以键入：

rowmarg=total(padgett rows)

与之类似，为了汇总整个数据中的全部值，可利用 total 命令：

bign=total(padgett)

如果我们研究两个行动者共同参与了哪些关系，则可以利用 Multiplex 命令。从技术角度讲，如果 $G(V, \{R_i\})$ 代表连接点集 V 和关系集合 $\{R_i\}$ 之间的一个多元关系图(multirelational graph)，并且点 v 和点 w 是 G 中的两个点，那么连接点 v 到点 w 的关系束(bundle of relations) B_{vw} 可界定为，$B_{vw}=\{R_i: vR_i w\}$。令 M_k 代表关系束的集合。多丛图(multiplex graph)就是一个多值图，其邻接矩阵的取值 $X_{i,j}=k$，k 是连接 i 到 j 的关系束的数目。简单地说，Transform→Multiplex 程序会计算出连接任何两个点的不同关系的类型数，并用一个数值表示。例如，两个矩阵以及行动者共同参与的关系类型属如下所示：

$$\begin{matrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{matrix} \quad + \quad \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{matrix} \quad \xrightarrow{\text{Multiplex}} \quad \begin{matrix} 0 & 1 & 2 & 3 \\ 2 & 0 & 1 & 1 \\ 1 & 0 & 0 & 2 \\ 3 & 1 & 0 & 0 \end{matrix}$$

该网络的成员会表现出四类关系组合：在两种关系上都无关系(第 0 类)；在两种关系上都有关系(第 1 类)；在第一种关系上无关系，但在第二种关系上有关系(第 2 类)；在第一种关系上有关系，但在第二种关系上无关系(第 3 类)。由上述两个数据得到一个多丛关系矩阵(multiplex matrix)，即箭头右端矩阵。

十、标准化处理

另一类数据转换程序是标准化(normalizing)，可以对整个数据或者某些行

或列进行标准化处理。例如，我们可能希望一个数据矩阵的每行都有同样的均值和标准差。特别是当矩阵的每行是根据不同的回答者收集来的时候，这种要求往往更加强烈。例如，如果要求被调查者评价自己家到其他行动者家庭之间距离的话，某些人可能以千米来回答，某些人可能以米来回答。在这种情况下，可以将数据进行标准化处理，从而便于比较。

Transform→Normalize 程序可用来标准化。可利用的标准很多，如针对均值、边缘值、标准差，或者欧氏距离、z-值等。每一类标准化标准都可以针对整个矩阵，或者针对某些行或列分析。例如，我们可以使每一行都有均值 0.00，标准差为 1.00，这就需要利用 z-值标准。

也可以利用 Matrix Algebra 程序进行标准化处理。例如，假设你有一个名为“trade”的 24×24 矩阵，你可能希望对它进行标准化处理，从而使得各列总和为 1。可以采取如下措施：在 UCINET 中，首先打开“trade”数据，然后沿着 Tools→Matrix algbra 路径，点击后在“Enter command here”处键入如下命令：

```
COLSUM=TOTAL(TRADE COLUMNS)
COLSUM24=FILL(COLSUM 24 24)
NTRADE=DIV(TRADE COLSUM24)
```

第一个命令将创建一个行向量(row vector)，它实际上是一个 1 行 24 列的矩阵，其中每个值都对应着初始矩阵的对应列的总和。下一个命令用 FILL 命令把行向量复制 24 次，获得一个 24×24 矩阵。最后一个命令把 trade 数据的每一项元素都除以该元素所在列的总和，达到了标准化的目的。

十一、改变模态

还有一类特殊的数据转换程序，涉及矩阵模态的转变(mode change)，例如从 2-模数据中导出 1-模数据，将一个属性向量转变为 1-模矩阵等。

在网络分析中，2-模数据常常以隶属数据形式出现，这种数据记载的是一群行动者参与一系列事件的情况。该数据一般把行动者看成行，事件看成列。我们可以从这种数据中计算出行动者—行动者矩阵(actor-by-actor matrix)，这只需要计算每一对行动者共同参与事件的数目即可。其结果生成一个“共同参与”(co-occurrence)或者“共享成员”(co-membership)矩阵，进而可以按照通常的数据加以分析。当然也可以计算出事件—事件矩阵(event-by-event matrix)，对于

每一对事件来说,该矩阵记录了同时参加这两个事件的行动者的数目,用它来测量行动者参与事件的重叠程度。

从一个二值 2-模矩阵中生成一个 1-模方阵的最简单的方法就是执行 Data→Affiliations(2-mode to 1-mode) 程序。该程序把一个 2-模矩阵看成是输入数据,计算每一对行动者(事件)共同参与(拥有)事件(或行动者)的数目。应该意识到,这种计算过程与根据一种初始的非标准化的相似性测度来测量行动者(或者事件)在事件(行动者)方面的相似性是一样的。要选择各种相似性测度需要执行 Tools→Similarities 程序。例如,对于二值数据来说,一种有趣的选择就是利用“完全匹配测度”(exact matches measure),该测度将计算两个行动者共同参加或者都不参加同样事件的次数。这样看来,在评价行动者之间关系的时候,不参加同一事件就派上了用场。对于多值数据来说,相关系数测度(the correlation measure)可能是一项好的选择。对于相异性测量来说(dissimilarity measure),如欧氏距离,可运行 Tools→Dissimilarities 程序。

还有一类特殊的模式转换,即从单一向量中生成一个 1-模矩阵。例如,一种网络研究可能收集到多个行动者之间的基本信息,如性别、年龄,收入等。问题是,这些属性与他们之间的朋友关系模式(“朋友关系”矩阵)具有怎样的关系?更具体地说,性别不同,朋友关系的模式是否不同?一种回答此类问题的方法是把这些属性数据转换为“网络”(networks),并利用 QAP 程序计算二者之间的相关系数。例如,考虑性别属性,我们假定它是一个向量 S,男性编码为 1,女性编码为 2。我们可以在 UCINET 中,利用 Data→Attribute to matrix 程序,将研究者自己事先已知的属性数据向量(如命名为“性别向量”)作为输入数据,从而可以创建一个行动者—行动者矩阵(actor-by-actor matrix),得到“性别异同”矩阵,它记录每一对行动者是否属于同一性别。也就是说,如果行动者 i 和行动者 j 的性别相同,我们就设定 $X(i, j)=1$,否则 $X(i, j)=0$。然后可以利用 QAP 来计算“性别异同”矩阵与“朋友关系”矩阵之间的相关程度。

对于一个量纲式的变量(metric variable),如收入,我们一般不探讨相同收入的行动者有没有成为朋友的趋向,但可能追问收入之差为多大的时候能预测朋友情况,在这种情况下,我们将设定 $X(i, j)$ 等于行动者 i 和行动者 j 之间收入之差的绝对值。

与之相反的程序(即从 1-模矩阵中生成一个 2-模矩阵)也可以执行。例如,Transform→Incidence 程序会把一个点—点邻接矩阵转换为一个 2-模的点—线

发生矩阵(node-by-line incidence matrix),其中各行代表点,各列代表线,如果点 i 隶属于线 j,则 X(i, j)=1。该矩阵可被用作一个指标矩阵(indicator matrix),可作进一步分析。

本章小结

本章介绍了社会网络分析软件 UCINET 的使用方法,主要介绍了数据的输入、处理和转换方法。小型的数据易于输入和整理,大型的数据输入则需要一定的技巧。UCINET 可以转换 Excel 数据,也可以直接对用数据语言形式保存的数据进行转换。本章还介绍社会网络的形式化表征方法:图和矩阵。介绍了矩阵的基本运算法则,其中矩阵之积是比较重要的运算。本章用 UCINET 中的矩阵算法展示了矩阵运算的过程及含义,这些算法能够计算现实生活中的关系模式,因而比较重要。

第四章

均衡性分析

在现实生活中存在着两个人之间、三个人之间以及多个人之间的关系。两个人之间的关系可以是单向或互惠的，三个人之间的关系可以是对抗或均衡的。假设你收入较少，却与一位富翁好友经常到饭店吃饭，你的好友每次都买单，你是否觉得这不“平衡”，这种关系不是“互惠”？再假设有三个人 A，B，C。A 喜欢 B，而 B 讨厌 C，那么 A 应该是喜欢 C，还是讨厌 C？这三个人之间的情感关系如何能够达到整体上“均衡”？一般情况下人们会认为，两个人之间的关系应该是“人不犯我，我不犯人；人若犯我，我必犯人”。三个人之间的关系很可能表现为“朋友的朋友是我的朋友，朋友的敌人是我的敌人，敌人的朋友是我的敌人，敌人的敌人是我的朋友”。当然，在现实生活中，敌人的敌人可能是朋友，也可能还是敌人。这里需要考察的问题是，社会关系在什么情况下出现“互惠”、“平衡”？

“互惠”这个概念意味着关系是相互的，互惠性是人类社会发展的一个重要推动因素。两个人之间的关系是否有互惠性受他们所处社会结构和文化结构的影响。例如，在中国，朋友关系讲究“互通有无，相互帮助”。不过，逢年过节的时候，成年人可以给好友的孩子一定的、短时间内“一去不返”的压岁钱，也应该给自己的年迈双亲一定数目的“养老金”，这个养老金可真的是“一去不返”的。当然，我们可以认为成年人每年给老人的养老金是对父母的“回报”，但是不能简单地认为回报“养育之恩”是“互惠性”的，因为“养育之恩”是伦理性的问题，不能进行简单的测量和计算。总体而言，不同类型的关系表现出不同的互惠性。互惠性不仅体现在个体之间，也体现在组织之间、国家之间。例如，两国之间只有互惠互利，才能保持关系的长久。

有关互惠的研究有一些命题，例如，物以类聚、人以群分。这个命题断言人们之间的属性的相似性带来关系的建立。不过，现实世界中的很多“联盟”者恰恰因为他们拥有不同的性质才联盟的。这表明，两个社会行动者之间的相同的属性和不同的属性都可以带来“关系”。属性相同，可情投意合；属性不同，则利

益互补。换句话说，属性的相同与不同对于关系的建立和维持起作用的方式是不同的。

其实，社会网络分析更关注的是三方的关系，因为三方关系才是社会的基础（刘军，2006b）。本章主要探讨的就是三方关系中的均衡问题，首先探讨符号三方关系，然后分析有向三方关系中的均衡问题。

第一节　符号图中的均衡研究

最早关于三方关系的研究当然首推齐美尔（Simmel, 1902），他指出，在三方关系中，第三方可能扮演渔利者、调停者、中介者等不同的角色。齐美尔关注的是三个行动者之间的关系，而在社会心理学的均衡研究中，最初探讨的是两个行动者针对一个事件的态度是否均衡的问题。

一、社会心理学中的均衡研究

最早关于三方关系中的均衡的研究首推社会心理学家海德（Heider, 1946, 1958）。他提出的认知平衡理论关注的是这样的问题，如果某个人 P(Person)和另外一个人 O(Other)关于某个对象或人 X(Object)的态度不一致，那么二人之间的关系会怎样？这个问题可画图表示，如图 4.1。

海德关注的是 P 和 O 这两个人在对待 X 的态度上是否具有一致性。如果简单地看待态度，可以将它仅分为两类：喜欢和不喜欢。

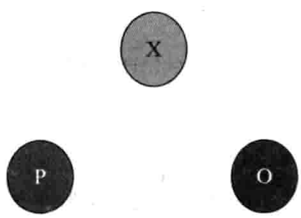


图 4.1　P-O-X 三方组

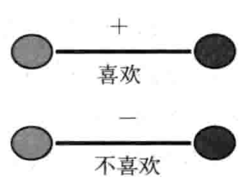


图 4.2　喜欢关系和不喜欢关系

如果考虑到 P-O-X 三方关系包含三对二方关系，即 P-O, O-X, X-P，再结合这两类态度，就可以形成八种三方关系（假定关系是对称的，即 A 如果喜欢 B，B 也喜欢 A），如图 4.3 所示。

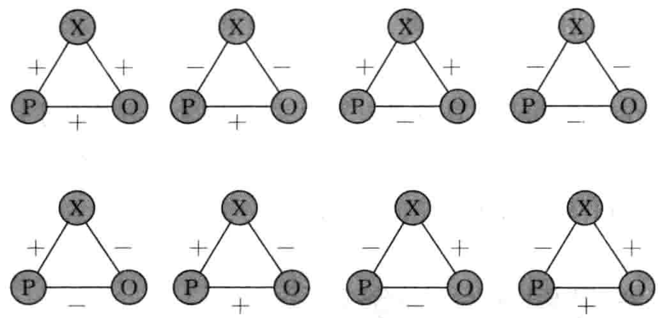


图 4.3　P-O-X 三方关系的八种表现

不过,如果不考虑具体的点,那么上述八类结构就可以化简为如图 4.4 的四个同构类。

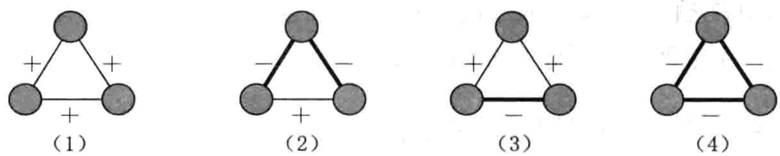


图 4.4　P-O-X 三方关系的四个同构类

问题在于,这几个同构类的结构能保持均衡吗?就这几个符号图来说,我们可以通过计算符号之积的方式来回答这个问题。符号之积为正者就是一个均衡图,符号之积为负者就是一个不均衡图。这就是海德的认知均衡理论的主要内容。当某人在某个话题上不赞同他的朋友时,此人会感到不舒服。

对于图 4.4(1)来讲,其符号之积为(+)(+)(+)=(+);这是一个均衡图,相当于说,"朋友的朋友就是朋友",三者之间的关系可以成为传递三方关系(transitive triad)。

对于图 4.4(2)来讲,其符号之积为(−)(−)(+)=(+);这也是一个均衡图,相当于说,"敌人的敌人就是朋友"。

对于图 4.4(3)来讲,其符号之积为(+)(+)(−)=(−);这是一个不均衡图,相当于说,"朋友的朋友就是敌人"。三者之间的关系可以成为非传递三方关系(intransitive triad)。

对于图 4.4(4)来讲,其符号之积为(−)(−)(−)=(−)。这是一个不均衡图,相当于说,"敌人的敌人还是敌人"。

接下来的问题就会出现:不均衡的结构在实际生活中会出现怎样的变化?按照海德的观点,随着时间的推移,不均衡的结构会逐渐走向均衡。如图 4.5

所示。

结构均衡可被视为一种基础的社会过程，可用于解释社会行动者之间的感情、态度、信念等是如何有助于促进稳定的(但未必无冲突)社会群体的生成的。不过，假定有两种互斥的情感状态(喜爱、讨厌)，同时假定相关的社会行动者坚持完全的理性行动，那么要想使得上述均衡状态出现，还必须有某种制度的保障才可以。其道理很简单：上述“均衡”原理所预设的法则与理性行动相悖。例如，如果交友原则是“我的敌人的朋友就是我的敌人”，这样的三方关系就会达到均衡。试想，一个“敌人”如果令自己坚守这个交友准则，那么他不会受到“眷顾”。可以想象，如果A和B相互视对方为敌人，那么A有很好的理由与B的朋友C以及B所有的朋友交朋友，这样就可以完全孤立B了(Kadushan，2012:23)。此时，A将敌人B的朋友变成自己的朋友，从而将B孤立起来，谁能说这不是一种均衡态？不过此时这三方关系符号之积可就是负的了。

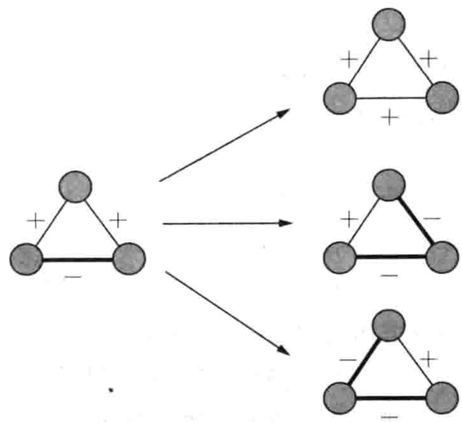


图 4.5　三方关系的演变

二、有关均衡的模拟研究

海德的均衡理论的基础比较简单，其中诸如信念、欲望、尝试和目的等意向性的概念框架居于核心地位。这些概念恰恰是心理学的“表面”领域和社会学的“深层”领域。人们对待他人的态度可以从“喜欢”转变为“讨厌”，这种改变的理论基础是海德提出的一些命题，即人们在自己的人际关系中趋向于选择均衡的状态，这也是由于人们在不均衡的状态中感觉到的压力带来的结果。海德的研究思辨性较多，均衡理论在20世纪50年代比较受关注，近年来却有些受冷落。不过，近年来有关均衡的模拟研究受到一定的重视。

对于三人以上的小群体来讲，海德(Heider，1946)的理论被卡特赖特等人(Cartwright & Harary，1956)和纽考姆(Newcomb，1961)推广，即将均衡理论推广到社会群体，得到代表人际关系网络的图的结构均衡理论(structural balance theory)。对于海德来说，小群体变迁的机制蕴含在行动者心中。而对于后几位学者来说，结构的均衡被视为群体层次过程的结果，即关注的是群体结

构。也就是说,有关均衡理论的文献大相径庭:一部分研究延续海德的思路,继续探讨行动者心灵深处的情感过程,关注二方组和三方组,此类研究只以海德的观念为基础,相关的研究结果和发现并不一致。另一部分研究则基于卡特赖特等学者的思路(Doreian, 2002),探讨群体层次的均衡过程。后一种思路只以卡特赖特和哈拉雷(Cartwright & Harary, 1956)的观念为基础,关心作为一个整体的群体结构,逐渐抛弃了海德基于人类群体的经验研究得到的洞见,其相关的研究结果也不一致。不过,行动者"内在的"动力机制可以通过沟通和传播,进而形成群体层次的社会结构,与此同时,群体层次的过程也可以生成宏观层次的结构,该结构既能够促成也可以限制微观层次的过程。

研究结果的不一致可能缘于这两种经验研究之间的互不借鉴。哈蒙等人(Hummon & Doreian,2003)"基于能动者的仿真模型"(agent-based simulation model),将二者结合提出了一种将微观机制和宏观机制结合的新的社会均衡论模型,该模型可以解释先前的文献中发现的不一致。他们的模型模拟了互依的社会行动者,令他们在达致均衡的认知状态过程中相互进行正向的和负向的选择。模拟中的设计变量包括群体规模、群体的敌对程度(操作化为负关系在初始条件中出现的概率)和行动者用什么模式选择既存的社会联系及其符号。选择的模式不同,行动者对自己所在的群体就会产生各种信息量的变化。在该动态模型中使用了群体层次的均衡机制,该机制的学理基础是由多瑞安等学者(Doreian & Mrvar, 1996)提出来的所谓"分区均衡观念"(the idea of partition balance):随着时间的推移,行动者的诸多选择会演变成这样的网络,即它将群体的成员分区成为稳定、均衡的子集合。包括群体规模等变量在内的几个设计变量对于行动者在达致均衡过程中作出的选择数、群体不均衡的层次、自认为网络是不均衡的行动者的数量以及所形成的子集合数量等都有复杂的影响。

他们的研究表明:(1)初始的敌对状态不可忽视;(2)沟通的模式对于均衡论的动态演变来说很重要;(3)相对于在"小型的"小群体来说,在"大型的"小群体中这些动态性有所不同;(4)在模型形成的子集合数量和实现的两类均衡之间存在着某种微妙的关系。这些结果既简单又具有挑衅性。这说明,人与人之间的符号网络往往趋于均衡,这个原本似乎常识性的、经验性的结构均衡论假设是否成立还真不好说。尽管经验研究都表明网络会走向均衡,但是很少有社会群体会达到完美的均衡。多瑞安等学者(Doreian & Krackhardt, 2001)认为,在既定的经验情景下可能存在着多个均衡机制。多瑞安(Doreian, 2002)又指出,如果说走向均衡的举动的确存在,那么它既不是简单的,也不是直接的。哈蒙等学者

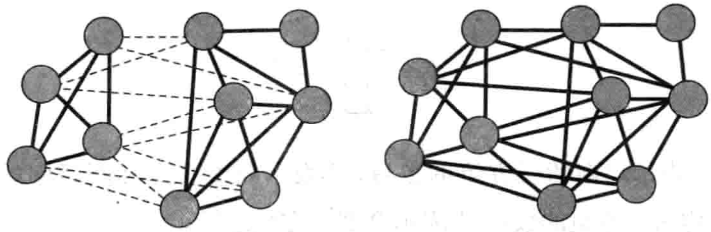
(Hummon & Doreian，2003:46)通过仿真研究，指出这里可能存在两个解释。第一，如果每个行动者都认为网络的图景是均衡的，他就无须改变其社会关系。如果超出三方结构，海德关注的内在性机制就不起作用。由此可知，尽管有关一个网络的认知图景都是均衡的，但是该网络的宏观层次结构却可以是不均衡的。第二个解释来自这样一个发现，即群体层次的均衡需要经历的步骤远大于个体层次均衡所需要的步骤。多瑞安等学者(Doreian et al.，1996)通过重新分析纽考姆的数据(Newcomb data)发现，互惠性、传递性和均衡性这三种理论机制起作用所历经的时间是不一样的，其中"均衡论"机制经历的时间最长。模拟研究也表明，均衡理论的机制在群体层次上其作用所经历的时间可能被低估了(Hummon & Doreian，2003:47)。这项研究也表明，基于能动者的仿真能够探究微观均衡过程和宏观结构过程及结果之间的关联，同时也能够考察宏观结构过程是如何限定微观均衡过程的。

三、整体均衡研究

在一个整体网络中，如果人人都追求这种均衡过程，那么所有的非传递性三方关系都会逐渐被消除，整个图会逐渐演变两个子图，出现结构上的均衡。这种分析将个人之间的情感关系视为人际关系网(Hummon & Doreian，2003；Kadushin，2012)。结构均衡理论认为，二人关系符号的改变能够带来整个图的均衡的变化，它可用点—线图表达。该理论可表述如下(Hummon & Doreian，2002)：

> 对于一个大群体中的某个图(多个个体构成的网络)来讲，当且仅当下面的条件成立，该网络就是均衡的网络，即该群体可以分为两个子群体，其中每个子群体内部的关系都是正的，不同子群体成员之间的关系都是负的。

所出现的两个子图就是"均衡的对抗图"和"完备图"，如图4.6所示。



均衡的对抗图　　　　完备图

注：实线表示"朋友关系"，虚线表示"敌对关系"。

图4.6　两类图

如果考察三个行动者之间的有向关系而不仅仅是符号关系,就可以研究三个人之间的“人际均衡”现象。如果人人都追求均衡,非传递性的三方关系会逐渐减少,整个网络会逐渐演变上述两个子图。不过,这主要是发生在符号关系中。

如果关注一个大的网络是如何变迁的,可以探讨其中的小结构(三方组、二方组)最初是如何变化的。因此,我们可以探讨包含的各种可能的符号三方组以及其中包含的二方组的变化。当然,这种思考方式的前提是,网络结构的“巨变”是由于其中小结构的“微变”带来的。所以,我们可以研究所有二方关系的初始符号模式,探讨当其变化之后,会给三方组带来什么变化(即三方组变“均衡”了,还是“不均衡”),再探讨随着时间的推移给整个网络结构带来怎样的巨变。

哈蒙和多瑞安(Hummon & Doreian, 2003)提出一种基于能动者的模型,其中的行动者在三方关系中为了达到均衡态,可以在某些分区的基础上改变他们的关系符号。与此同时,有学者(Wang & Thorngate, 2003)尝试通过使三方组随机地变成均衡,进而观察一个网络是如何一分为两个子群体的。还有学者(Khanafiah & Situngkir, 2011)在这两项研究的基础上,探讨二方关系层次上的情感关系变化是如何影响到人际关系网络中的整体均衡指数的。

为了进行这样的分析,有必要给出一个“整体均衡指数”(global balance index),可将它界定为“均衡三方组的数量与全体三方组的数量之比”。还可以界定“局部均衡指数”(local balance index),界定为“均衡三方组数与由每一个二方组形成的三方组数之比”。用这两个指数可以测量一个网络的(局部或整体)均衡程度。如果一个网络在整体上是均衡的,其整体均衡指数就是1,即表明该网络就整体而言是均衡的网络(Khanafiah & Situngkir, 2011)。

整体均衡指数可以表述为:

$$\beta=\frac{\sum\limits_{J\leqslant I}T_{balanced}}{\sum\limits_{I}T_{tot}}$$

其中 $T_{balanced}$ 表示均衡的三方组的数目,T 表示在整个人际网络中的三方组总数,J 表示均衡三方组的数量,I 表示全部三方组总量。

根据均衡指数就可以确定一种系统性的反馈,换句话说,如果一个情感关系的符号改变了,整个人际关系系统就会考量是接受还是拒绝符号的变化。不过

这里就需要有一个前提假定，即假定一个网络会逐渐走向高层次的均衡。也就是说，如果情感关系符号的改变导致均衡指数更高，则接受这种改变。

在由某种情感关系（“+”关系、“-”关系或无关系）联系的规模为 N 的群体中，其二方关系总量为：

$$N=\frac{N!}{(N-2)!2!}$$

如果所形成的二方关系类型是三类（不管是正关系，负关系还是无关系），那么可能存在的关系模式为：

$$p=3^D=3^{\left(\frac{N!}{(N-2)!2!}\right)}$$

在由 N 个行动者组成的群体中，三方组的各种可能组合的数目便是：

$$T_{tot}=\frac{N!}{(N-3)!3!}$$

例如，对于 4 人群体来讲，各种可能的二人之间的关系数量为 $C_4^2=\frac{4!}{2!2!}=$ 6 个（即 1-2，1-3，1-4，2-3，2-4，3-4 这六种）。如果情感关系有三类（即喜欢，用+号表示；不喜欢，用-号表示；无关系，用 0 表示），那么来自这四人群体的各种可能的情感关系数等于 $3^6=729$ 种关系模式。同理，四人群体中的三人关系数，即三人组数量为 $C_4^3=\frac{4!}{3!3!}=4$ 个（1-2-3；1-2-4；1-3-4；2-3-4）。

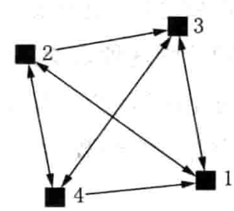


图 4.7　四人关系

我们可用一个 $N\times N$ 邻接矩阵探讨关系的模式，如表 4.1 所示。

表 4.1　规模为 4 的群体中的各种三方组组合

三方组的组合			形成的关系		
1	2	3	12	13	23
1	2	4	12	14	24
1	3	4	12	14	34
2	3	4	23	24	34

根据这个矩阵可以确定其中的三方组是否均衡,然后可以计算网络的均衡指数。例如,我们可以随机地改变初始矩阵的元素,测量其均衡指数。符号从 0 变成 1 或 −1 的情况很可能出现,因为在一个群体中,人们都有动力去相互结识。而符号从 +1 变成 0 的情况则极少见,因为人们几乎不可能忘记先前结识的朋友。随着关系符号的改变,针对每一次变动便会得到新的均衡值数。如果新的均衡值数达到了更高层次的均衡,则接受这种变动。

有两位学者(Khanafiah & Situngkir, 2011)利用模拟法探讨了上述模型。第一次模拟采用规模为 8 的网络,先将其初始矩阵进行随机化处理,然后再进行 350 次迭代,结果发现该群体达到了均衡态,其均衡指数从 0.0714 提高到 1。此时该人际关系网将不再改变,因为均衡值数达到了最大。研究发现,网络达到均衡状态所经历的时间很短(大约 87 次迭代),也验证了社会有丝分裂理论(social mitosis theory)(Wang & Thorngate, 2003),随着时间的推移,网络中有一个行动者从均衡的群体中分离出去,剩下的大群体完全是由正的人际关系构成。

当然,这种分析及模型只适用于相对较小的行动者群体。为了使得模型更有现实性,应该再考虑局部均衡指数这个概念。与上述模型不同,加入局部均衡指数之后,模型要考虑的因素是,二方关系的改变如果能够提高局部均衡指数,则接受这种改变。局部均衡指数可以界定为“由一个关系的改变带来的均衡三方组的数量”与“由该关系的改变带来的所有可能的三方组数”(而不是网络中的全部三方组数)之比。可用公式表示如下:

$$\beta_{local}^{x,y}=\frac{\sum\limits_{J_{x,y}\leqslant I_{x,y}} T_{balanced}}{\sum\limits_{I_{x,y}} T_{local}}$$

其中 $T_{balanced}$ 表示由 x-y 这个二人组形成的均衡三方组数量,其中 T_{local} 表示由 x-y 二人组形成的三方组总数。在由该二方组形成的局部三方组中,我们关注 J,即均衡三方组的数量,I 是对应形成的三方组总量。例如,对于规模为 50 个点的网络来讲,其中包含二方组 1 225 个,三方组 19 600 个。如果某个二方关系的改变会带来三方组的局部均衡指数的增加(不考虑是否带来整体均衡指数),则接受这种改变。

我们可以考察局部三方组的均衡过程对网络中全体三方组的均衡有什么影响。特别是在大网络中更应该考察这一点,因为在大网络中,个体的信息有限,

不大可能认识全体人员并与之保持均衡的关系。因此，单独的个体只能考虑其局部的关系，坚持有限理性的原则行事，他获得的有关均衡条件的信息取决于其在网络中占据的局部位置。基于这样的思考可以建立模拟的模型，考察局部的均衡出现的步骤。模拟的结果发现，该大型网络通过 16 000 次迭代会平稳地达到均衡态（Khanafiah & Situngkir，2011：9）。

总之，这两种模型模拟的结果表明，整个网络达致均衡所需要的时间不同。前一种模型适用于发现小型网络中的均衡态，后一种模型适用于发现大型网络中的均衡态。不管怎样，社会均衡理论都为我们提供一种分析社会系统中的情感关系如何演化并达到均衡的工具。

需要指出的是，上述研究过程及结果需要研究者建立特定的模拟程序，UCINET 并不提供此类程序。

最后需要补充的是，结构均衡理论虽然可应用于分析符号网络中的块模型结构，不过，大多数现实中的符号网络并不是完全均衡的。原因之一可能在于，人们很少探究符号社会网的生成动力机制，不关注符号网络的过程及结果。例如，可能有的行动者与一对相互之间有消极关系的行动者都保持着积极关系；这一对行动者可能隶属于敌对的群体之中；某些行动者可能被其网络成员认为拥有积极的关系，但是实际表现出来的关系却可能是消极的，也有可能某些行动者表面一套，背后一套。如果认为这些现象不符合传统的均衡原理，但是为什么它们大量存在呢？既然它们大量存在，就应该有其合理性。换句话说，我们可以将这些看似违背结构均衡原理的现象看成是“合理的”，不过它们合乎的未必是“均衡”之理，可能合乎其他互动之“理”，例如人们在作出积极或消极举动的时候坚持“调和”之理、遵守“差序声望”（differential popularity）的原则，在自己的群体内反倒因各种原因而出现“群内敌对性”（internal subgroup hostility）。通过对这些观念的形式化分析，多瑞安等学者（Doreian & Mevar，2009）推广了结构均衡块模型，引出了“松散的结构均衡模型”（relaxed structural balance blockmodel）。他们用实际数据对这种松散的模型进行了检验，表明它更能描述现实世界中符号网络的均衡关系。该模型及其实现过程可通过另外一种网络分析软件 Pajek 实现，UCINET 本身不能执行计算，因此这里从略。感兴趣的读者可阅读沃特·德·诺伊等学者的文章（Wouter de Nooy，2012：121）。

第二节　有向关系中的均衡

海德提出的理论主要是“认知均衡”理论，关注的是两个人针对某个特定事物(或第三个人)的认知而形成的符号性的三方关系。最初，该理论关注的关系都是对称的。但是，现实中的关系更多的是不对称的关系，对称的关系并不多(至少就情感关系来讲是这样)。这样的话，我们有必要探讨在有向关系中的均衡问题，首先考察有方向的符号关系。

一、有向 P-O-X 三角关系

诸如情感、喜欢等关系往往是有方向的，A 喜欢 B，但是 B 未必喜欢 A。因此，均衡理论可以拓展到有方向的标记图中，如图 4.8 所示。

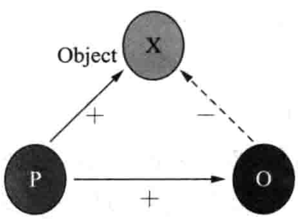


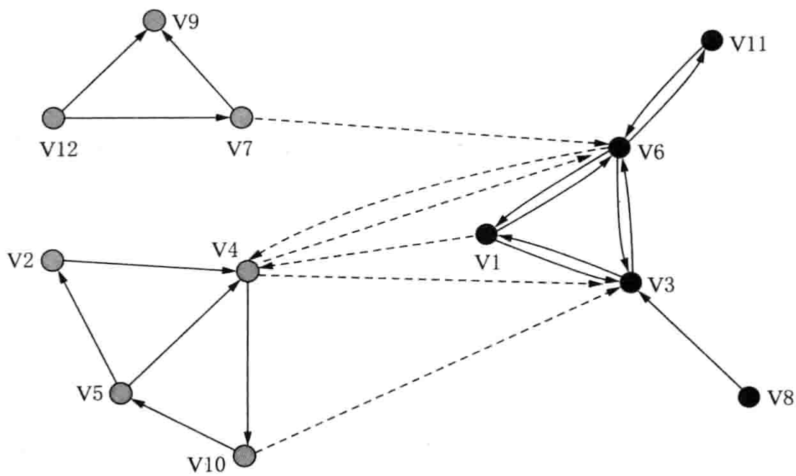
图 4.8　P-O-X 有向三方组

在图 4.8 中，从 P 到 O 到 X 再回到 P 可形成一个循环，不过由于线的方向与循环的序列走向不一致，所以这是一个“弱循环”。这个弱循环是不均衡的，因为其中有奇数条负号的线。换句话说，对于循环或弱循环来讲，只有其包含的负号的线有偶数条的时候，该循环才是均衡的(沃特・德・诺伊等，2012:114)。海德关注的是感觉关系，比如 P 对 O 的看法以及 P 对待事件 X 的看法等。不过，即便 O 对 X 的看法也是用 P 的感受来测量的，未必与 O 的真实感受一致，这就是社会心理学中的“归因”现象。

有向图可以达到均衡，其结果分成两个聚类，在聚类内部的关系都是正的，在聚类之间的关系都是负的。当一个有向标记图中所有的弱循环都是平衡的时候，该图就是平衡的。

不过，现实中一个网络未必分成两个聚类，可能分成三个或更多个聚类。多个聚类之间的关系可以是均衡的，这意味着平衡性这个概念应该拓展为可聚类性(clusterability)。具体地讲，对于一个网络来说，如果它满足下面的条件，则称之为可聚类的：该网络可分为多个聚类，正连线连接的都是同一聚类的点，负连线连接的都是不同聚类的点。研究证明，如果网络中不存在只含单条负线的弱循环，该网络就是可聚类的。可见，均衡性只是可聚类性的一个特例。

图 4.9 展示了一个平衡的图。在该图中分为三个聚类，正关系出现在聚类内部，负关系出现在聚类之间，这样的网络才可以达到均衡。



资料来源：沃特·德·诺伊等，2012：116。

图 4.9　一个平衡图

在现实生活中，一个群体分成两个均衡的对抗群体的情况并不多，分成多个均衡群体的情况倒不少，这是否表明均衡理论不成立呢？当然不是。现实的多个子群体的存在要求我们允许所有的负号的三方关系的存在。也就是说，我们将(一)(一)(一)并不看成是不存在的，而看成是存在的。换句话说，敌人之间的敌对关系可以存在，可以不像好友关系那样有激发和团结的力量。

二、正—负关系的均衡性分析

在上文介绍的 P-O-X 关系中，X 项表示的主要是事项或事件。如果 X 表示个人，P-O-X 就变成三个人之间的符号关系。这里所说的"符号"可以是正号和负号，正号可以表示"喜欢"、"支持"、"服从"等关系，负号可以表示"不喜欢"、"反对"、"对抗"等关系。由此引出一个可以研究的问题：在一个整体网络的任何三个人 A，B，C 之间，所有可能的关系类型有多少种？具体地说，A 喜欢 B、B 喜欢 C，同时 A 也喜欢 C 的三方组情况有多少？或者说，A 喜欢 B、B 既喜欢又讨厌 C，而 A 讨厌 C 的三方组情况又有多少？诸如此类的三方组很多。下面用一个例子加以展示。

假设在六个人之间存在两类关系,一类是喜欢关系,如下面的数据矩阵所示,对应的数据文件命名为 6positive;另一类是"不喜欢"关系,具体关系如下面的不喜欢关系矩阵所示,对应的数据文件命名为 6negative。还假设已知 1, 2, 3 这三个行动者为第一小组,4, 5, 6 为第二小组,用一个列向量数字表示,前三个值是 1,后三个值是 2,该列向量命名为 6partition。

表 4.2 "喜欢关系"矩阵

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	1.00	1.00	0.00	1.00	0.00
2	1.00	0.00	1.00	0.00	1.00	0.00
3	1.00	1.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	1.00	1.00
5	0.00	1.00	0.00	1.00	0.00	1.00
6	0.00	0.00	0.00	1.00	1.00	0.00

表 4.3 "不喜欢关系"矩阵

	1	2	3	4	5	6
	----	----	----	----	----	----
1	0.00	0.00	1.00	1.00	1.00	1.00
2	0.00	0.00	0.00	1.00	1.00	1.00
3	0.00	0.00	0.00	0.00	1.00	1.00
4	1.00	1.00	0.00	0.00	0.00	0.00
5	1.00	1.00	0.00	0.00	0.00	1.00
6	0.00	0.00	1.00	0.00	0.00	0.00

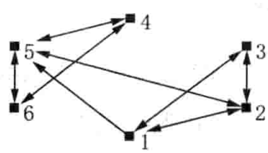


图 4.10 "喜欢"关系图

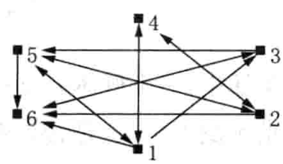


图 4.11 "不喜欢"关系图

UCINET 中,依次点击 Network→Balance Counter,即出现计算符号三方关系的对话框。在"Positive Network:"项目上键入"6positive",在"Negative Networks:"上键入"6negative",在"Attribute Dataset:"项目中键入"6partition",并且在"Which Column:"右侧键入"1"。点击 OK 后,该程序就会计算出各类三方组的统计量来,这样的符号三方组共有 216 种。

计算结果中有"网络层次的统计量",它包含每一类三方组的数量。计算结果还有"点层次的统计量",它是一个矩阵,其中的行表示三方组类型,列表示各个点,第 i, j 项给出的是包含点 j 的第 i 类三方组的数量。

下面给出的是网络层次的统计量。

表 4.4 网络层次"均衡分析"结果

```
BALANCE COUNTER
---------------------------------------
Triad Frequencies
iRj|jRk|iRk          Freq         Prop
-------------   -----------  -----------
    Ps|Ps|Ps            6        0.018
    Ps|Ps|Nd            0        0.000
    ...
    Ps|Ns|Ns            9        0.027
     Ps|Ns|s            6        0.018
...
     Ps|s|Ns            5        0.015
      Ps|s|s            2        0.006
......
    Ns|Ps|Ps            2        0.006
...
    Ns|Ns|Ps           10        0.030
    Ns|Ns|Pd            0        0.000
     Ns|Ns|s            1        0.003
...
  Ps|Ps|PNs             2        0.006
 Ps|PNs|PNs             2        0.006
  Ns|Ns|PNs             4        0.012
...
    s|Ps|PNs            1        0.003
...
    s|PNs|Ps            2        0.006
    Ps|PNs|s            1        0.003
```

在这个结果中,Ps 表示处于同一个分区中两个行动者之间的积极关系,Pd 表示处于不同分区中两个行动者之间的积极关系,Ns 表示处于同一个分区中两个行动者之间的消极关系,Nd 表示处于不同分区中两个行动者之间的消极关系,s 表示在同一个分区中两个行动者之间无关系的情况,d 表示在不同分区中

两个行动者之间无关系的情况。

由此可见,第一行表示在“处于同一个分区中两个行动者之间的积极关系”(Ps|Ps|Ps)中,满足“i 指向 j、j 指向 k,并且 i 再指向 k”(iRj|jRk|iRk)这个条件的关系数量有 6 个,占总关系数的 1.8%(0.018)。

同理,计算结果的中间有一项为“Ps|PNs|PNs 2 0.006”,它的含义是:在满足“i 指向 j、j 指向 k,并且 i 再指向 k”(iRj|jRk|iRk)这个条件的三方关系中,“i 和 j 处于同一个分区(s 的含义),并且 i 喜欢 j,而 j 既喜欢又不喜欢 k(PNs),并且 i 也既喜欢又不喜欢 k(PNs)”,满足这几个条件的三方组有两个,占总数的 6‰(0.006)。

下面给出的是点层次的统计量。

表 4.5 点层次均衡分析结果

Transpose

	1	2	3	4	5	6
1 Ps|Ps|Ps	4	3	2	2	2	2
2 Ps|Ps|Pd	0	0	0	0	0	0
3 Ps|Ps|Ns	2	2	2	1	1	0
...						
13 Ps|Ns|Ps	1	2	0	0	3	0
14 Ps|Ns|Pd	0	0	0	0	0	0
15 Ps|Ns|Ns	3	3	2	1	1	0
...						
25 Ps|s|Ps	1	1	0	0	1	0
26 Ps|s|Pd	0	0	0	0	0	0
27 Ps|s|Ns	1	1	0	2	1	2
...						
73 Ns|Ps|Ps	2	2	1	2	1	0
74 Ns|Ps|Pd	0	0	0	0	0	0
75 Ns|Ps|Ns	2	2	2	2	0	0
...						
97 Ns|s|Ps	2	2	2	0	0	1
98 Ns|s|Pd	0	0	0	0	0	0
...						

157	s|Ns|Ps	0	0	2	2	1	2
…							
169	s|s|Ps	0	0	0	0	1	0
…							
216	d|d|d	0	0	0	0	0	0

第一行结果表明，在满足“Ps|Ps|Ps”(即在处于同一个分区的三个行动者 i, j, k 中，如果 i 喜欢 j，j 喜欢 k，那么 i 就喜欢 k，这实际上是“传递三方组”)这个条件的三方组中，行动者 1 参与了四个这样的三方组，行动者 2 参与了三个，其余四个行动者都参与了两个。据此大致可以认为，第 1 个行动者和第 2 个行动者在“传递三方组”上表现得更积极一些。对其余结果的解释与此类似。

三、有向三方组的谱系

在现实生活中，三个人(而不是两个人对待同一个事件)之间的关系也存在均衡问题。因此，我们应超出符号关系，分析有方向的三方关系的结构。为了分析三方关系，首先要分析二方关系，因为二方关系是三方关系的基础。就二方关系来讲，其有向的关系只有如下三类：

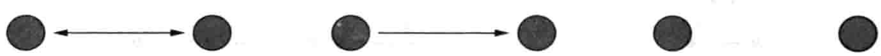


图 4.12　二方关系的三个同构类

二方关系有三类，而每一个三方关系又都是由三个二方关系构成的。这样的话，我们可以根据三方关系中包含的每一类二方关系的数量区分出三方关系的类型。为方便起见，将区分出来的三方关系进行命名，采用的命名系统为 MAN 标签系统(M 表示 Mutual dyad，即互惠对，A 表示 Asymmetric dyad，即不对称对，N 表示 Null dyad，虚无对)。

这个性质是三方关系的一种性质，它用于描述的关系是有向关系。而对于有向的三方关系(triad)来说，其所有可能存在的状态有 64 个，其中有 16 个同构类(见图 4.13，图中略去了“虚无三人组”003，即三个点之间都没有任何连线)(Holland & Leinhardt, 1970)。

为了描述这 16 个同构类，霍兰等学者提出如下标记规则，每个类别用三个或四个记号表征：

第一个数字给出三方组中互惠对的个数；

第二个数字给出三方组中不对称对的个数；

第三个数字给出三方组中虚无对的个数；

最后一个字母(如果需要的话)对相似的两个三方组进行区分。

例如,有两个 030 三方组,它们很相似,都包含三个不对称对。为了区分起见,其后分别用“T” 和“C”标记于后,即 030T 和 030C。其中,“T”代表这是一个“传递关系”(transitivity),“C”代表这是一个“循环关系”(cycle)。还有两个记号的含义是,“D”代表“向下”(Down)的关系,“U”代表“向上”(Up)的关系。

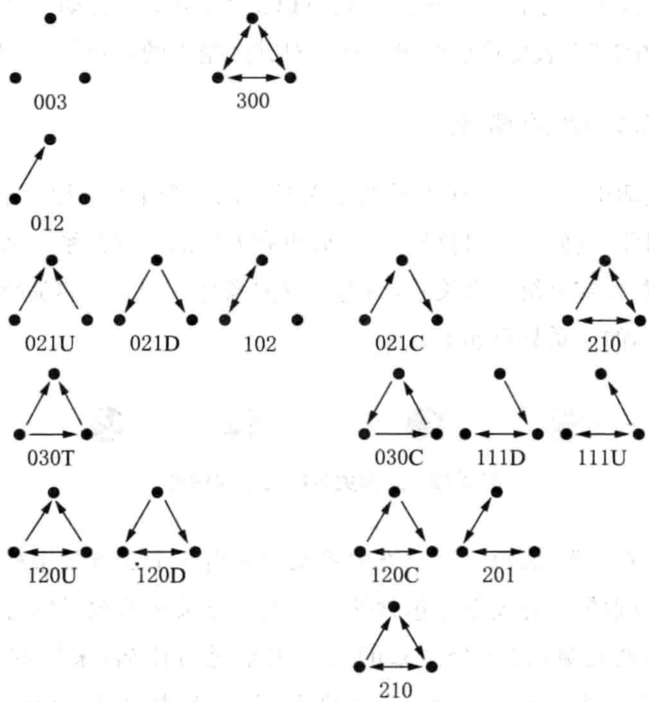


图 4.13 三方组的 16 个同构类

在一个拥有 g 个点的有向网络中可能存在的三方组共有组合数 C_g^3 个,其中的每一个必然与上述 16 个类别之一同构。如果用 T_u 表示属于第 u 个同构类的三方组的数目,就可以用一个包含 16 个取值的列向量 T 表示所有可能的 16 个同构类的数目。学者们约定的列向量为:

$$T=(T_{003}, T_{012}, T_{102}, T_{021D}, T_{021U}, T_{021C}, T_{111D}, T_{111U}, T_{030T}, T_{030C}, T_{201}, T_{120D}, T_{120U}, T_{120C}, T_{210}, T_{300})'$$

这个列向量就叫做三方谱(triad census),如图 4.13 所示。其所有值的总和等于 C_g^3。

三方谱是有向图的一系列图的性质的汇总,它抓住了一些重要的结构特征(Wasserman & Faust, 1994:569)。例如,当人们遇到困境的时候,通常希望自己的朋友能够帮忙或得到朋友的朋友帮助。这表明人们是希望这种朋友关系能够有“传递”。可以用描述性的和形式化的语言表示如下。如果行动者 i 有指向 j 的关系,j 有指向 k 的关系,那么我们期望 i 有指向 k 的关系。换句话说,如果 i→j 并且 j→k,那么 i→k。满足这个性质的三方关系称为传递三方关系(transitive triad),它与上文介绍的符号三方组中的传递关系类似。如果行动者 i 有指向 j 的关系,j 有指向 k 的关系,k 又有指向 i 的关系,即 i→j, j→k, k→i,满足这个性质的三方关系称为循环三方关系,这也是一类重要的三方组。

三方组谱系分析尤其适用于分析大型的网络结构。如果研究的兴趣只是分析小群体中的点,那么三方关系分析未必提供什么帮助。如果探讨图中存在多少孤立点,三方谱中也不提供这个信息。

以下举例说明利用 UCINET 计算三方谱的过程。笔者有一个关于 103 个家庭之间的“紧密关系”数据(参见刘军,2006),文件名为 den。在 UCINET 中,点击 Network→Triad Census 命令,在出现的对话框中选出要分析的数据文件 den,再点击 OK,即可得到其三方谱如下。

表 4.6　三方谱系分析结果

```
TRIAD CENSUS
--------------------------------------------------------
Input Network dataset:        den(D:\datafiles\den)
Output Measures:              den-Tri(D:\Data\den-Tri)

Triad Census for dataset den
                         den
                   ---------
 1    003        129296.00
 2    012             0.00
 3    102         43126.00
 4   021D             0.00
 5   021U             0.00
 6   021C             0.00
 7   111D             0.00
 8   111U             0.00
 9   030T             0.00
```

10	030C	0.00
11	201	3691.00
12	120D	0.00
13	120U	0.00
14	120C	0.00
15	210	0.00
16	300	738.00

可见,该数据中包含的 102 同构类有 43 126 个,300 类有 738 个。如果要分析这些值是否显著,需要知道三方谱分布的性质,特别是其均值和方差,具体计算参见沃瑟曼和福斯特(Wasserman & Faust, 1994:577)。

三方谱研究具有一定的理论意义。通过分析三方谱,可以把整个网络还原为 16 个汇总的统计量,分析其各种结构性质,包括结构均衡的性质。三方组中的某些结构界定了某种行为机制的存在。例如,021C、111D、111U、201、030C 这五个三方组都包含非传递性;300、120D、120U、030T 这四个三方组都包含传递性;030T 更是一种完全的等级结构。如果一个网络中的全部三方组都是这样的结构(如图 4.14 所示),则该网络就是完全的等级结构网络。120C 和 210 则是混合的三方组。003 是虚无三方组,即任何两个点之间都不存在联系。021D 是 B 有两个关系,一个指向 A,另一个指向 C,因此这是出-2 星(out-star)结构,它表示关系都向外发出。021U 则表示入-2-星,是得到关系的结构。聚类结构有 003, 300, 102 等。其他三方组的结构及其意义不再赘述。

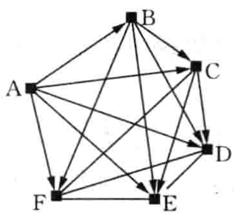


图 4.14　完全等级结构网

我们可以利用一个网络的三方组分布检验假定存在的机制是否成立。这就需要在该网络中首先计算每一种三方组的数量,将该值与假定网络是随机情况下的期望值比较,从而对某种假设进行检验。总之,通过分析这些三方组在整个网络中存在的数量,可以揭示网络的结构性质是否具有传递性等特征。因此,网络越大,对三方谱的分析越有价值,这种分析能够揭示宏观社会结构的微观基础,也是研究网络动态演变的基础,这些都体现了三方谱研究的理论意义。

在什么意义上说三方谱分析的结果能够表现出"结构均衡"的性质?这个问题取决于一些因素,如均衡的含义、所分析的关系的性质、三方谱的结构等。例如,如果研究的关系是"借钱关系",那么出-2-星结构(即 A→B, A→C)就不是均衡的,但却体现了中心者 A 的核心地位,混-2-星结构(即 A→B, B→C)和一个互

惠对(即 A→B, B→A)可以看成是均衡的。而如果研究的关系是“给面子”的关系,那么出-2-星结构(即 A→B, A→C)还是不均衡的,不过,混-2-星结构不应视为均衡的,但是一个互惠对可以看成是均衡的,因为“给面子”行为主要发生在两个人之间。另外,均衡还是一个动态的过程。例如,在两个间点收集“建议关系”数据,得到两个数据矩阵。如果在第一个矩阵中有 A→B,没有 B→A,而在后一个矩阵中出现了 B→A,我就可以说这两个矩阵合在一起,才在 A 和 B 之间表现出均衡来。换句话说,如果在一个时间点矩阵中体现了不均衡,在另外一个时间点矩阵中可能“找回”均衡,在局部体现了不均衡,在整体上可能体现了均衡。

本 章 小 结

海德开创了社会心理学领域中符号网络的均衡研究,提出了人际关系的均衡原理:网络中消极关系数量为偶数时,整个网络为均衡的。卡特赖特将海德的研究推广到群体,他探讨了群体层次的均衡,提出整体网络的结构均衡原理。近年来的仿真研究揭示了均衡过程的机制,提出了整体均衡指数和局部均衡指数。在有向整体网络中,三方关系有 16 个同构类,对它们的谱系分析可以揭示整体网络中的均衡性、互惠性、传递性等结构。

第五章

中心性和影响力*

“权力”(power)是社会科学中的一个重要概念。什么是“权力”,不同的学科和学者给出了不同的回答。在社会网络学者看来,一个抽象的社会行动者是没有权力的。一个社会行动者之所以拥有权力,是因为他与他者存在关系,可以影响他人。或者说,一个人的权力就是他者对此人有依赖性。

社会网学者从“关系”角度出发对权力进行定量研究,基于“中心性”给出多种关于权力的量化指标,即中心度和中心势指数,这可以看成是社会网络分析对权力研究的独特贡献。

“中心性”是社会网络分析的研究重点之一。个人或者组织在其社会网络中具有怎样的权力,或者说居于怎样的中心地位,这一思想是社会网络分析者最早探讨的内容之一。在最初的研究中,点的中心度(centrality of a point)和图的“中心势”(centralization of a graph)这两个概念常常混淆不清。例如,在弗里曼(Freeman, 1979)的研究工作中,他既谈到了“点的中心度”(point centrality),又谈到了“图的中心度”(graph centrality)。斯科特(Scott, 2013:83)指出,如果令“中心度”这一术语特指为点的中心度,“中心势”特指作为一个整体的图的中心度,那么,所有的混淆都将消失。可见,中心度测量的是个体在整个网络中的权力,“中心势”测量的是一个图在多大程度上围绕某个或某些特殊点建构起来。

在研究中心度和中心势的时候,一般坚持这样的思路。首先给出一个点的各种“绝对中心度”的表达式;然后,为了比较来自不同图的点的中心度,需要给出“相对中心度”指数,即“标准化”的绝对中心度指数。计算一个点的“相对中心度”指数的原则是,该点的“绝对中心度”除以图中其他点的最大可能的中心度。最后,给出一个图在整体上的中心势指数。

以下结合美国加州大学尔湾分校弗里曼教授的经典文章(Freeman, 1979)

---

* 本章是在刘军(2004:第五章)的基础上增补而成的。

以及有关文献，介绍比较常用的几类中心度以及中心势指数，包括：度数中心度、中间中心度、接近中心度以及与之相应的中心势指数。最后介绍针对子群体的中心度分析。

第一节　度数中心性

前面已指出，根据的标准不同，用以刻画中心度的指标也不同。

一、点的度数中心度

度数中心度是一个较简单的指数。行动者 x 的度数中心度(point centrality)可以分为两类：绝对中心度和相对中心度。

(一) 绝对度数中心度

点 A 的度数中心度就是与点 A 直接相连的其他点的个数。如果一个点与许多点直接相连，我们就说该点具有较高的度数中心度。

如果用 C_{AD} 代表绝对度数中心度，那么，一个点 x 的绝对度数中心度的表达式为 $C_{AD}(x)$。如果某点具有最高的度数，则称该点居于网络的中心，拥有权力。在测量度数中心度的时候根据的是与该点直接相连的点数，并不考虑间接相连的点。因此，所测量出来的中心度可以称为“局部中心度”(local centrality)。在有向图中，每个点的度数可分为点入度(in-degree centrality)和点出度(out-degree centrality)。一个点的点入度是进入到该点的其他点的个数，即该点得到的直接关系数。一个点的点出度是该点直接发出的关系数。

(二) 相对度数中心度

当图的规模不同的时候，不同图中点的局部中心度不可比较。为了比较不同图中点的度数中心度的大小，弗里曼(Freeman, 1979)提出了相对度数中心度：点的绝对中心度与图中点的最大可能的度数之比。在一个规模为 n 点的图中，任何一点的最大可能的度数一定是 $n-1$。因此，在一个有 10 个点的图中，度数为 6 就意味着相对中心度为 $6/(10-1)=0.66$。相对中心度也可用于比较、测量有向图中点的中心度，其具体表示方法参见刘军(2004：119—120)。如果网络是有方向的，其中一点 x 的相对点度中心度(记作 RD)的表达式为：$C'_{RD}(x)=$(x 的点入度＋x 的点出度)/$(2n-2)$，其中，n 是网络的规模。

二、图的度数中心势

除了探讨图中某点在多大程度上居于核心之外,还可以探讨一个图在多大程度上表现出向某个点集中的趋势,这就需要研究图的中心势。例如,在星形网络图中,“核心点”的度数中心度最大,其他点的度数中心度都是1,这种网络中点的度数中心度差异很大,所有其他点都向核心点集中。正是在这个意义上,我们说该图具有大的中心势。又如,在n点完备网络中,任何点的度数都等于n－1,该图就没有表现出中心趋势,正是在这个意义上,我们说该图的中心势为0。为了定量地探讨图的中心趋势,需要设计出一种中心势指数,用以刻画整个图的这种中心势,也可以达到比较不同图的中心趋势的目的。前文已经指出,为了避免混淆,我们用中心度来描述图中任何一点在网络中占据的核心性;用中心势刻画网络图的整体中心性。

图的中心势指数的构造来源于如下思想。首先找到图中各个点的最大中心度的值;然后计算该值与图中其他点的中心度的差,从而得到多个“差值”;再计算这些“差值”的总和;最后用这个总和除以在理论上各个差值总和的最大可能值。用公式表示如下:

$$C=\frac{\sum_{i=1}^{n}(C_{\max}-C_i)}{\max[\sum_{i=1}^{n}(C_{\max}-C_i)]}$$

在具体计算的时候,我们既可以利用点i的绝对中心度(记为C_{ADi}),也可以利用其相对中心度(记为C_{RDi})。如果图中点的绝对中心度的最大值记为$C_{AD\max}$,相对中心度的最大值记为$C_{RD\max}$,那么上述公式的具体形式为:

$$C_{AD}=\frac{\sum_{i=1}^{n}(C_{AD\max}-C_{ADi})}{\max[\sum_{i=1}^{n}(C_{AD\max}-C_{ADi})]}$$

问题是,在中心势的定义中,“理论上各个差值总和的最大可能值”是多少?其实,对于规模为n的网络来讲,只有当它是星形网络(star network)这种“理想的”网络的时候,分母才达到“理论上的各个差值总和的最大可能值”,其具体取值为n^2-3n+2,因此,上述公式又可以进一步变为:

$$C_{AD} = \frac{\sum_{i=1}^{n}(C_{AD\max} - C_{ADi})}{n^2 - 3n + 2}$$

如果根据相对中心度计算，那么把分子除以$(n-1)$就得到相对中心度。因此，分子和分母同时除以$(n-1)$，上述公式就变为“相对度数中心势指数”如下：

$$C_{RD} = \frac{\sum_{i=1}^{n}(C_{RD\max} - C_{RDi})}{n - 2}$$

这两个公式是相等的，可以在 UCINET 中计算出来。

在计算度数中心度和度数中心势的时候，根据的是“直接关系”，不考虑间接关系，或者说不考虑整个网络中的全部关系，只关注点的局部关系。因而，可以将这两个指数看成是局部的中心性指数。

第二节　中间中心性

一、点的中间中心度

（一）中间中心度的含义

直觉告诉我们，如果很多人要想建立联系必须要经过某个人，则可以认为此人居于重要地位，因为“处于这种位置的个人可以通过控制或者曲解信息的传递而影响群体”(Freeman，1979:221)。因此，另一个刻画行动者个体中心度的指标是中间中心度(betweenness centrality)，它测量的是行动者对资源控制的程度。具体地说，如果一个点处于许多其他点对(pair of nodes)的捷径(geodesic)(最短的途径)上，就说该点具有较高的中间中心度。

要测量一个点的中间中心度，首先要了解“中间性”(betweenness)这个概念。中间性主要是由美国社会学家林顿·弗里曼(Freeman，1979)教授提出来的一个概念，它测量的是一个点在多大程度上位于图中其他“点对”的“中间”。他认为，如果一个行动者处于多对行动者之间，那么他的度数一般较低。这个相对来说度数比较低的点可能起到重要的“中介”作用，因而处于网络的中心。根据这个思路就可以测量点的中间中心度。

（二）中间中心度的测量

假设在一对点(pair of points)X 和 Z 之间存在多条捷径，其中有一条可能

经过点 Y。点 Y 相对于点对 X 和 Z 的中间度(betweenness)指的是该点处于此对点的捷径上的能力,具体利用“中间性比例”(betweenness proportion)这个概念来刻画这种“能力”,其定义为:经过点 Y 并且连接这两点的捷径数与这两点之间的捷径总数之比。它测量的是 Y 在多大程度上位于 X 和 Z 的“中间”,这就是“中间中心度”(具体参见刘军,2004:122—126)。

具体地说,假设点 j 和点 k 之间存在的捷径条数用 g_{jk} 来表示。点 j 和点 k 之间存在的经过第三个点 i 的捷径数目用 $g_{jk}(i)$ 来表示。第三个点 i 能够控制此两点的交往的能力用 $b_{jk}(i)$ 来表示,它等于 i 处于点 j 和点 k 之间的捷径上的概率,即 $b_{jk}(i)=g_{jk}(i)/g_{jk}$。

把点 i 相应于图中全部点对的中间度加在一起,就得到该点的绝对中间中心度(记为 C_{ABi}),$C_{ABi}=\sum_{j}^{n}\sum_{k}^{n}b_{jk}(i)$,$j\neq k\neq i$,并且 $j<k$。

我们还可以给出标准化的中间中心度,即相对中间中心度。只有在星形网络情况下,图中点的中间中心度 C_{ABi} 才可能达到如下最大值 $C_{max}=(n^2-3n+2)/2$(证明过程参见刘军,2004:124)。因此,点 i 的相对中间中心度(C_{RBi})为 $C_{RBi}=\frac{2C_{ABi}}{n^2-3n+2}$,其取值范围为 0 和 1 之间,并且该值可用于比较不同网络图中点的中间中心度。

点的中间中心度测量的是该点在多大程度上控制他人之间的交往。如果一个点的中间中心度为 0,意味着该点不能控制任何行动者,处于网络的边缘;如果一个点的中间中心度为 1,意味着该点可以 100% 地控制其他行动者,它处于网络的核心,拥有很大的权力。关于中间中心度指数的研究还有很多新的进展,具体参见本书第九章。

二、图的中间中心势

与度数中心势指数的计算类似,我们也可以计算中间中心势指数。对于规模为 n 的图来讲,首先找到图中各个点的中间中心度的最大值;然后计算该值与图中其他点的中间中心度之差,从而得到多个“差值”;再计算这些“差值”的总和;最后用这个总和除以在理论上该差值总和的最大可能值(在规模为 n 的星形网络中,该差值总和才达到最大值 n^3-4n^2+5n-2)。具体而言,一个图的中间中心势指数可表达为:

$$C_B = \frac{\sum_{i=1}^{n}(C_{AB\max} - C_{ABi})}{n^3 - 4n^2 + 5n - 2} = \frac{\sum_{i=1}^{n}(C_{RB\max} - C_{RBi})}{n-1}$$

其中，$C_{AB\max}$是规模为 n 的图中点的绝对中间中心度的最大值，$C_{RB\max}$是其对应的点的相对中间中心度。星形网络具有 100%的中间中心势指数，环形网络的中间中心势指数为 0。

三、线的中间中心度

除了考察点的中间中心度以外，还可以分析线(或边)的中间中心度。令 b_{jk} 是连接点 j 和点 k，并且经过边 i 的捷径与连接点 j 和点 k 的捷径总数之比。边 i 的中间度便是各个 b_{jk} 的总和(j 和 k 是不同的点)。可见，线的中间中心度测量的是一条线出现在一条捷径上的次数。该指标测量的是一条线对信息的控制程度。在测量一条线 i 的中间中心度时，需要考虑网络中所有经过 i 的两点之间的捷径，计算 i 在全部捷径中所占比例。

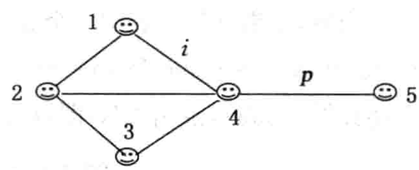


图 5.1　一个 5 点图

以图 5.1 为例，我们分析连接点 1 和点 4 的线 i 的中间中心度，这就要考虑所有经过 i 的捷径。首先分析行动者 1 和行动者 4，会发现 1-4 之间只有一条捷径，即 i。因此，i 对于点 1 和点 4 来说能够 100%地控制二者的信息，或者说 i 相对于点 1 和点 4 来说的中间中心度是 1。再看点 1 和点 3。这两个点之间存在两条捷径，其中有一半(即 1 条)经过 i。因此，i 能够 50%地控制 1 和 3。也就是说，i 相对于点 1 和点 3 这对点来说其线的中间中心度为 0.5。点 1 和点 2 的捷径不经过 i，所以，i 相对于点 1 和点 2 这对点来说的线中间中心度为 0。最后看点 1 和点 5，二者的捷径只有一条，并且经过 i，所以，i 相对于点 1 和点 5 这对点来说的线中间中心度为 1。这些值的总和为 1＋0.5＋1＝2.5。因此，线 i 的中间中心度是 2.5。同理可以分析得到点 4 和点 5 之间的线 p 的中间中心度是 4(请读者自行分析)。

线的总"中间中心度"尽管算起来比较复杂，但是这一测度仍然具有直观意义，即"控制信息"。

线的中间中心度可以在 UCINET 中轻松地算出来。例如，将上述 5 点图矩阵输入 UCINET，并命名为 5nodes。在 UCINET 中，沿着 Network→Centrality

and Power→Freeman Betweeness→Edge(line) Betweenness,选择上述数据5nodes,点击 OK,即可得到如下结果:

表 5.1 5 点图的边中间度分析

Edge Betweenness

		1 1	2 2	3 3	4 4	5 5
1	1	0.000	1.500	0.000	2.500	0.000
2	2	1.500	0.000	1.500	2.000	0.000
3	3	0.000	1.500	0.000	2.500	0.000
4	4	2.500	2.000	2.500	0.000	4.000
5	5	0.000	0.000	0.000	4.000	0.000

可见,点 1 和点 2 之间的线的中间中心度为 1.5,点 1 和点 4 之间的线的中间中心度为 2.5,点 2 和点 3 之间的线的中间中心度为 1.5。其他点之间的线的中间中心度都可以从表中看出。一共可以分析出 6 条线的中间中心度。

当然,如果计算上表中 5 个点的中间中心度,需要在 UCINET 中沿着 Network→Centrality and Power→Freeman Betweeness→Node Betweenness,选择上述数据 5nodes,点击 OK,即可得到如下结果:

表 5.2 5 点图的点中间度分析

Un-normalized centralization: 13.500

		1 Betweenness	2 nBetweenness
4	4	3.500	58.333
2	2	0.500	8.333
1	1	0.000	0.000
3	3	0.000	0.000
5	5	0.000	0.000

Network Centralization Index = 56.25 %

可见,点 1 的中间中心度为 0,点 2 的中间中心度为 0.5,点 3、点 4、点 5 的中间中心度分别是 0、3.5 和 0。整个网络的标准化中间中心势为 0.562 5,这个值是比较大的,表明整个网络表现出比较明显的向某个点集中的趋势。

点的中间中心度与线的中间中心度不同。在一个网络中,有多少个点,就存

在多少个点的中间中心度，有多少条线，就存在多少个线的中间中心度。点的中间中心度测量的是单独的一个行动者的控制优势，线的中间中心度测量的是两个点之间的关系在整个网络中居于怎样的控制优势，二者有所不同。

四、针对中间中心度的等级嵌套分析

有时候我们关注这样一个问题：在一个图中，能否先找到中间中心度最大或者说居于最核心的一些点，再找到中间中心度稍小一些的其他点？这种想法是可以实现的。在 UCINET 中，可利用基于中间中心度的等级简化程序（Hierarchical Reduction）进行分析。该程序能够根据各个点的中间中心度，生成由各个点构成的一个嵌套系列。

在具体计算过程中，首先还是要计算出每个点的中间中心度，将中间中心度是 0 的点删除，针对删除后的简化图再计算各个点的中间中心度，然后再删除中间中心度是 0 的点，如此反复，直到所有的点都被删除，由此得到一个嵌套系列。最初的时候将全部点都纳入等级之中，然后删除中间中心度是 0 的点。该程序可针对有向图进行分析。

下面，我们结合黑龙江省一个村落——法村的帮工关系中的一个子群体数据（有关该数据的具体情况，参见刘军，2006a：238—248，391）进行这种嵌套分析。在 UCINET 中，沿着 Network→Centrality and Power→Freeman Betweeness→Hierarchical Reduction，选择“帮工第一子群”数据，点击 OK 得到如下结果：

表 5.3 帮工第一子群的中间中心度嵌套分析

```
SUCCESSIVE NODE ELIMINATION BY BETWEENNESS
----------------------------------------------------
Input dataset:      H:\Dissertation.all\data\帮工第一子群
                              1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
            1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
            - - - - - - - - - - - - - - - - - - - - - - - - -
          1 3 3 3 3 3 3 3 3 3 3 3 3 1 2 1 3 3 3 3 3 1 3 3 3 3
```

这个表的每一列都对应一个点，一共有 25 列，对应着帮工子群中的 25 个家庭。最下面一行给出了每一次删除哪些点的情况。在第 i 列的值 k 意味着村民 i 在第 k 次迭代中被删除。从中可以看出，第一次删除的点有 13，15，21；第二次删除的点有 14；其余各个点就是第三次删除的点。这表明，点 13，点 15，点 21 的中间中心度最低，其次是点 14，最后剩下的点的中间中心度比较高。这一点

也体现在下面的简化表中：

表 5.4　帮工第一子群的中间中心度嵌套发生阵

```
Successive Reduction of H:\Dissertation.all\data\帮工第一子群.##h via Betweenness

            1 1 2 1           1 1 1       1 1 1 1 2   2 2 2 2
            3 5 1 4 4 5 7 8 6 0 1 2 1 2 3 6 7 8 9 0 9 2 3 4 5
            1 1 2 1 4 5 7 8 6 1 1 1 1 2 3 1 1 1 1 2 9 2 2 2 2
            - - - - - - - - - - - - - - - - - - - - - - - - -
  3       3 . . . . 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  2       2 . . . 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  1       1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

上述分区发生阵中最下面一行的值都是“1”，这说明程序在第一次计算的时候将全部行动者都拿进来。在矩阵中的第二行，可以看出 13，15 和 21 号村民家庭的中间中心度较低。

第三节　接近中心性

除了上述介绍的度数中心度、中间中心度之外，还有一些刻画点的权力的指标。可以想象，一个点越是与其他点接近，该点在传递信息方面就更加容易，因而可能居于网络的中心。这就是“接近中心性”(也叫整体中心性)思想。

一、点的接近中心度

(一) 接近中心度的含义

点 x 的接近中心度(closeness centrality)是一种针对不受他人控制的测度。弗里曼等学者根据点与点之间的“距离”(distances)来测量“接近中心度”。如果一个点与网络中所有其他点的“距离”都很短，则称该点具有较高的整体中心度(又叫做接近中心度)。在图中，这样的点与许多其他点都“接近”。当我们的研究不需要对直接关系进行考察的时候，接近中心度就是一个有用的概念。

(二) 接近中心度的测量

一个点的接近中心度是该点与图中所有其他点的捷径距离之和(sum of distances)。其表达式如下所示：

$$C_{APi}^{-1} = \sum_{j=1}^{n} d_{ij}$$

其中 d_{ij} 是点 i 和点 j 之间的捷径距离(即捷径中包含的线数)。

实际上，这是“绝对接近中心度”。为了对来自不同图中的点的接近中心度进行比较，需要给出“相对接近中心度”指数。

只有在星形网络中，C_{APi}^{-1}才可能达到最小值。对于包含 n 个点的星形网络来说，“核心点”的接近中心度是 $n-1$（证明参见刘军，2004：128）。“绝对接近中心度”除以这个最小的接近中心度，就可以得到“相对接近中心度”（即接近性程度的标准化指标）为 $C_{RPi}^{-1}=\frac{C_{APi}^{-1}}{n-1}$ 或者 $C_{RPi}=\frac{n-1}{C_{APi}^{-1}}$。

这个相对值容易解释，并且可以用来比较来自规模不同的网络中的两个点的接近中心度的大小。根据接近中心度的含义可以看出，与中心点距离最远的行动者也在信息资源、权力、声望以及影响方面最弱。注意，接近中心度的值越大，越说明该点不是网络的核心点。

二、图的接近中心势

一个图的接近中心势指数（closeness centralization）表达式为（具体推导过程参见 Freeman，1979：231）：

$$C_C=\frac{\sum_{i=1}^{n}(C'_{RC\max}-C'_{RCi})}{(n-2)(n-1)}(2n-3)$$

与度数中心度类似的是，星形网络具有 100% 的接近集中趋势，而完备网络、环形网络的接近中心势为 0。

一般来讲，上述三种中心度是相关的。如果它们不相关（或相关系数较小），可能表明网络具有令人感兴趣的结构。三者之间可能存在的关系如表 5.5 所示。

表 5.5　三类中心度之间的关系

	度数中心度低	接近中心度低	中间中心度低
度数中心度高		所嵌入的聚类远离网络的其他点	“自我”的联络人是绕过他的冗余的交往关系
接近中心度高	是与重要人物有关联的关键人物		在网络中可能存在多条途径，自我与很多点都接近，但是其他点与另外一些点也很近
中间中心度高	“自我”的少数关系对于网络流动来说至关重要	此类点极少见。意味着“自我”垄断了从少数人指向很多人的关系	

以上探讨了三类中心度及中心势的定义和测度。问题在于,在实际操作过程中,到底应该选择哪种指数进行测量?这里只能给出一些原则性的说明。首先要理解这三种中心度测度的含义是不同的。度数中心度测量的是一个点与其他点发展直接交往关系的能力。接近中心度和中间中心度刻画的是一个点控制网络中其他行动者之间交往的能力,它依赖于行动者与网络中的所有行动者之间的关系,而不仅仅是与邻点之间的直接关系。有时候,接近中心度测量的结果没有中间中心度测量的结果那样精确。但是,总的来说,三种中心度测量的结果相差不大。弗里曼也指出,到底选哪个指标?"这依赖于研究问题的背景,如果关注交往活动,可采用以度数为基础的测度;如果研究对交往的控制,可利用中间中心度;如果分析相对于信息传递的独立性或者有效性,可采用接近中心度。不管怎样,对于上述三种测度来说,星形网络的中心点都居于核心地位。"(Freeman, 1979:226)

最后一点需要指出的是,在测量的时候,上述三类指标可能产生不一致的结果,并且它们也没有考虑到行动者之间的交换或者交往的规模,因而仅适用于对二值网络的测量。

三、多种中心度指数的同时计算

由于研究的目的,我们可能需要同时计算出各个点的多种中心度指数,进而可以作对比分析。在 UCINET 中,这一工作可利用 Network→Centrality and Power→Multiple Measures 来计算。利用该算法可计算出四种标准化的中心度测度,即度数中心度、接近中心度、中间中心度和特征向量中心度。具体的计算过程及结果从略。但需要指出的是,该算法并没有计算出各种描述统计量,也没有计算出中心势指数。

第四节 影响力分析

在一个整体网络中,每个行动者的关系选择是不同的。有的行动者明显得到很多人的关照,体现在他得到的关系选项多,即他的影响力大。有的行动者虽然不被关注,即得到的关系选项少,但是他可能选择较多的他人,即发出的关系多。当然,如果一个行动者得到的关系和发出的关系都少,他便是"孤苦伶仃"之人了,不会有影响力。这只是对"影响力"的简单分析和解说。其实,在网络中,

"影响力"的含义比这里的解说要复杂一些。本节介绍几个影响力指数,并用实例加以讲解。

一、影响力的含义及测量

对影响力可以有多个角度进行测量。下面介绍最简单的一种,如何分析其缺陷,再给出比较合理的影响力指数。

(一)常规的影响力指数

对社会影响的研究由来已久。通常情况下,人们用"得到他人关注"的次数来测量"影响",这就是常识意义上的影响力指数。用矩阵语言表述,矩阵中"列总和"大的行动者的影响力就大,地位就高。例如,假设研究六个人之间的"支持"关系,其数据矩阵如下所示,在 UCINET 中保存并命名为 6 KATZ。

表 5.6 6 点关系矩阵

	A	B	C	D	E	F
	----	----	----	----	----	----
A	0.00	0.00	0.00	0.00	0.00	1.00
B	0.00	0.00	1.00	0.00	0.00	1.00
C	0.00	1.00	0.00	1.00	0.00	1.00
D	1.00	0.00	0.00	0.00	1.00	0.00
E	0.00	0.00	0.00	1.00	0.00	1.00
F	1.00	0.00	0.00	1.00	0.00	0.00
列总和	2.00	1.00	1.00	3.00	1.00	4.00

现在,记各个列总和为 s_j。由于该矩阵是二值矩阵,其规模是 $n=6$,其中每个点的度数最大为 $n-6=5$。因此,列总和 s_j 除以 $n-1=5$,就得到常规的地位指数:

$$常规的地位指数向量=(0.4,\ 0.2,\ 0.2,\ 0.6,\ 0.2,\ 0.8)$$

一般来讲,我们的确是根据一个人 X 得到直接关系的多少来判断其影响力的,这种指标关注的是"多少人选择了 X"。不过,如果不追问"多少人选择了 X",而是深究"谁选择了 X",就会发现该网络的一些重要性质,这些性质表明常规的影响力指数有缺陷。

首先,F 和 D 的地位显然最高。然而,A 却被 F 和 D 这两个地位高的人选择,尽管 A 再也没有被其他人选择。难道 A 的地位不高于常规计算的结果?

其次,地位低的三个人的位置并不对等。B 和 C 互选对方,再无他选。而 E 通过 D 与群体的其他人有联系,其位置是不同于 B 和 C 的。另外,F 选择了 D,但是没有得到回选。

这些性质表明,常规的影响力指数有缺陷,最主要的原因在于它只考虑直接的关系,忽视了间接的关系。

(二) 考虑到了间接关系的影响力指数

要想考虑到间接关系,就需要考量矩阵的幂次。关系矩阵的各个幂次之所以可用来测量"影响",是因为它们计算了任何一对点之间的各种长度的途径的数量。下面结合文献(Katz, 1953; Hubbell, 1965; Taylor, 1969),介绍具体的计算方法。首先介绍卡兹建构的影响力指数(Katz, 1953)。

1. 卡兹的影响力指数。

令上述 6 点矩阵为 C。其中的元素 c_{ij} 表示行动者 j 得到 i 的"选择"或"支持"。前文已经指出,C 平方中的元素 $c_{ij}^{(2)}$ 表示 i 和 j 两点之间存在的长度为 2 的途径数目,即行动者 i 经历多少条长度为 2 的途径才能够到达 j。换句话说,$c_{ij}^{(2)}$ 表示从 i 指向 j 的"长度为 2 的途径的条数"。同理,$c_{ij}^{(3)}$ 表示从 i 指向 j 的"长度为 3 的途径的条数"。C 的更高次幂中元素的含义与之类似。矩阵 C 的列总和表示对应的点得到矩阵中其他点的选择的总次数。矩阵二次方 C^2 的列总和表示对应的点得到矩阵中其他点经历两步选择的总次数,三次方 C^3 的列总和的含义与之类似。将这些矩阵相加,就可以构造一个影响指数。考虑到矩阵的幂次越高,其元素的对于构造地位指数的效应越低。因此,有必要引入一个"衰减因子"刻画这种表象。

不过,在构造影响指数之前需要给出一些前提性的假设,这些假定也符合建构一般的模型或指数时坚持的思维逻辑。首先要假定两个行动者之间只要存在直接或间接的"联系",就认为二者之间有"影响";没有联系,便无沟通,何谈"影响"? 其次,每个"联系"都相互独立,并且对"影响"带来效应的可能性也相同。虽然这个假设显然没有前一个假设更"真",但它也是一个接近"真"的合理的假设。这就要求我们给出一个常数 a,它依赖于所研究的群体及情境。这样的话,如果存在一条"联系"的长度为 k,即存在一条 k-链(k-chain),那么其对"影响"产生效应的可能性即为 a^k。这里的 a 测量的是链条中的非一衰减度(non-attenuation),$a=0$ 表示完全衰减,$a=1$ 表明无衰减。在这个模型中,C、C^2、C^3 等矩

阵的列总和对应的权重分别是 a、a^2、a^3 等。

前文指出，a 的取值取决于群体及情境，下面稍加详述。假设我们关注群体中某种信息的传递。可以想象，针对同样的信息，不同群体会有不同的反应和传递模式。例如，在一个村落中，村民在年末会得到国家的“补助”，如现金、米面等。那么，一个村落会得到多少袋大米？村落中不同的群体会对这个信息有不同的传递模式。例如，“村干部”群体未必希望全体村民都知道具体的“补助”量，低收入村民则希望知道补助的细节，包括分配方案。同一个群体针对不同的信息也有不同的反应。例如，对于村落中某个家族来讲，该家族内部的“私密”信息不会外传，而其他家庭的丑闻可能会积极传播开来。这说明，对于处于特定情境下的某个群体中的各种长度的链条来说，a 可能是常数，也可能不是。如果 a 不是常数，便有必要考察比较复杂的模型。出于建模的方便，一般要假设 a 是一致的常数。

令 s_j 表示矩阵 C 的第 j 列的总和，s 为各个要素 s_j 构成的一个列向量。例如，在上例中，转置后的行向量 $s'=(2, 1, 1, 3, 1, 4)$。按照上述说明，我们希望发现下面这个“和矩阵”的列总和。

$$T=aC+a^2C^2+\cdots+a^kC^k+\cdots=\sum_{k=1}^{\infty}a^kC^k=(I-aC)^{-1}-I$$

矩阵 T 的元素为 t_{ij}，其列总和为 $t_j=\sum_i t_{ij}$。令 t 表示由该矩阵各个列总和元素 t_j 构成的列向量，u 表示一个其元素都是 1 的列向量，二者的列数相等。于是便有 $t'=u'[(I-aC)^{-1}-I]$。在该等式两端右乘 $(I-aC)$，即可得到：

$$t'(I-aC)=u'-u'(I-aC)=au'C$$

对两端实施转置，得到：

$$(I-aC')t=aC'u$$

但是，$C'u$ 是一个列向量，其元素为 C' 的各个行总和，当然也是 C 的各个列总和。因此，$C'u=s$。最后，将等式两端都除以 a，得到：

$$\left(\frac{1}{a}I-C'\right)t=s$$

如此看来,给定了 a、C、s 之后,只需解这个线性方程组即可得到 t,$t=\left(\frac{1}{a}I-C'\right)^{-1}s$,而无须计算矩阵 C 的每个幂次。

当 $1/a$ 的值介于矩阵 C 的最大特征值和二倍的最大特征值之间的时候,计算结果符合实际意义。如果 $1/a$ 的的值较小,长链对于要构造的"影响指数"的效应就大。为了计算的方便,通常取 $1/a$ 的值为整数。可以取接近矩阵的最大特征值的那个整数值。例如,在上述 6 点图矩阵中,最大的特征值小于 1.7,因此取 $1/a$ 的值为 2。对于其中的元素都非负的一个矩阵来说,其最大特征值的一个简单的上限就是最大的行总和(或列总和)。

用列总和 s_j 除以 $n-1$(最多可能的选项数),就得到常规的地位指数。利用同样的标记,可以得到 t_j 的被除数 $(n-1)^{(k)}=(n-1)(n-2)\cdots(n-k)$,同时可以计算出 $m=a(n-1)+a^2(n-1)^{(2)}+a^3(n-1)^{(3)}+\cdots\approx(n-1)!\ a^{n-1}e^{1/a}$。

上述计算得到的 $t=\left(\frac{1}{a}I-C'\right)^{-1}s$ 就是影响指数。不过,计算 t/m 可以得到标准化的影响指数。在 UCINET 中提供的是 t 的值。

对于上述例子来讲,现在计算上述构造的地位指数。为简便起见,可令 $a=1/2$。按照方程 $\left(\frac{1}{a}I-C'\right)t=s$,令 C 的转置的相反数乘以影响系数 t_1, t_2, …, t_6 等,并将其对角线上的值加上 $1/a=2$,于是得到下面的方程组:

$$
\begin{array}{cccccccc}
2t_1 & & & -t_4 & & -t_6 & =2 \\
 & 2t_2 & -t_3 & & & & =1 \\
 & -t_2 & +2t_3 & & & & =1 \\
 & & -t_3 & +2t_4 & -t_5 & -t_6 & =3 \\
 & & & -t_4 & +2t_5 & & =1 \\
-t_1 & -t_2 & -t_3 & & -t_5 & +2t_6 & =4
\end{array}
$$

解方程,得到影响系数 t_1, t_2, …, t_6 分别为 13,1,1,11.4,6.2 和 12.6。尽管 $n=6$,但计算得到的估计值 $m=(6-1)!\ (0.5)^{6-1}e^{1/a}=27.71$ 与其实际值 26.25 却很接近。然后,将 t_i 都除以 27.71,得到标准化的影响系数如下:

新的影响系数向量=(0.47,0.04,0.04,0.41,0.22,0.45)

将这个新指标与上述常规指标比较可以发现有差异，也表明新指标能够更好地表明每个点的真正地位。

2. 胡贝尔的影响力指数。

卡兹(Katz, 1953)的影响力模型要求实现知道矩阵的特征值，据此确定衰减系数，这往往为计算影响力指数带来麻烦。另外，他提出的指数仅考虑到关系的“选择者”，没有考虑“被选者”，在计算一系列矩阵时没有考虑 $(aC)^0 = I$，而这一项可以反映系统外部因素对行动者之间“影响力”的贡献。所以，胡贝尔(Hubbell, 1965)将它加入计算之中给出了影响力指数，这种思路将矩阵看成是一个开放的系统，考虑到了“输入的影响”和“输出的影响”，因而更符合实际(Hubbell, 1965:381)。所以，在胡贝尔给出的影响力指数中，上述 T 矩阵变为：

$$T=(aC)^0+(aC)^1+a^2C^2+\cdots+a^kC^k+\cdots=I+\sum_{k=1}^{\infty}a^kC^k=(I-aC)^{-1}$$

$$t'=u'[(I-aC)^{-1}]$$

然后据此计算得到 t 的值，再像卡兹指数一样计算出影响指数。

3. 泰勒的影响力指数。

泰勒(Taylor, 1969:499)对卡兹指数进行了标准化处理，给出标准化的影响力指数。具体地说，泰勒针对卡兹提出的影响矩阵中的各个幂次，用行总和减去列总和，然后用该长度的总途径数进行标准化处理。用公式表示如下：

$$d_i^{(n)}=\frac{r_i^{(n)}-c_i^{(n)}}{\sum_{j=1}^{N}r_j^{(n)}}$$

这里的 $r_i^{(n)}$ 是矩阵 C 的第 n 次幂的第 i 个行总和，$c_i^{(n)}$ 是矩阵 C 的第 n 次幂的第 i 个列总和，二者之差有明确的意义：以距离 n 受到第 i 个行动者影响的人数减去第 i 个行动者受到影响的人数，显然，如果该值大于零，说明第 i 个行动者更容易影响他人而不是受到影响；如果该值小于零，说明第 i 个行动者更容易受到他人影响而不是影响他人。$\sum_{j=1}^{N}r_j^{(n)}$ 是矩阵 C 的第 n 次幂的全部行总和(当然它也等于全部列总和之和)。

在卡兹指数中，影响矩阵是 $P=(I-aC)^{-1}-I$，因此，针对该矩阵计算标准

化的值即可,具体的公式为 $d_i=\frac{r_i-c_i}{\sum_j r_j}$。这里的 r_i 是矩阵 P 的第 i 个行总和,c_i 矩阵 P 的第 i 个列总和。据此计算,即可得到标准化的影响指数。

二、影响力指数计算示例

在 UCINET 中,依次点击 Network→Centrality and Power→Hubbell/Katz Influence,出现的对话框程序可以计算整体网络中任何一对点之间的影响力测度。在对话框中,输入数据应选择已知的多值矩阵或邻接矩阵,计算方法可选上述三项,即胡贝尔指数、卡兹指数和泰勒指数。

如上文所述,胡贝尔在计算影响指数的时候,在矩阵序列中加入了单位矩阵(identity matrix),而卡兹指数则不用加入单位矩阵。对于胡贝尔指数来说,其影响矩阵是 $I+\sum_{k=1}^{\infty}a^kC^k$,它在极限特定情况下等于 $(I-aC)^{-1}$,这里的 C 是邻接矩阵。在卡兹指数中,影响矩阵是 $\sum_{k=1}^{\infty}a^kC^k$,在极限特定情况下等于 $(I-aC)^{-1}-I$ 的逆矩阵。

前文指出,一般要假定长的途径对"影响"的贡献小。所以当考虑全部途径总和时,要加入一个衰减因子。衰减因子(Attenuation Factor)β(Beta)通常设定为 0.5 即可。因为这个值应该小于主特征值的绝对值的倒数。之所以 β 设定为 0.5 即可,是因为数据矩阵的所有特征值都必然小于最大的行(或列)总和。在具体计算的时候,还涉及是否将矩阵的值都除以矩阵全部元素之和,默认的情况是不除。如果将初始矩阵的元素都除以矩阵全部元素之和,能够保障矩阵之幂结果收敛。计算结果给出的是影响矩阵(influence matrix),其中的第 i 行第 j 列给出的是行动者 i 对 j 的影响。

在 Hubbell/Katz Influence 对话框的 Input dataset 框右侧选出所保存的数据文件 6KATZ,在 Computational Method 项目选择 Katz, Attenuation Factor (Beta)项目键入 0.5, Divide matrix by overall sum? 项目选择 No,具体的对话框如图 5.2 所示。点击 OK 后会得到卡兹影响指数计算结果。然后,再回到该对话框,在 Computational Method 项目中点击三角标记,选择 HUBBELL,其余的设定不便,点击 OK 后会得到胡贝尔影响指数计算结果。

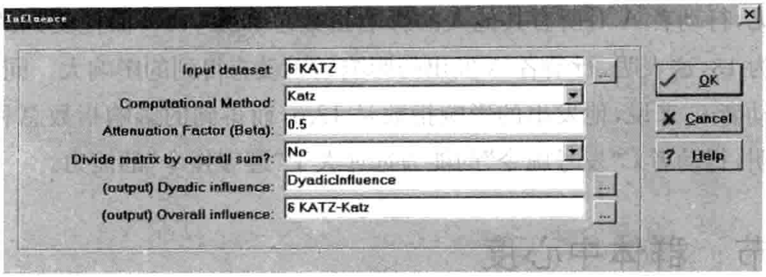


图 5.2　影响力分析对话框

这两个结果比较如下：

表 5.7　卡兹点对之间影响指数

		1 A	2 B	3 C	4 D	5 E	6 F
1	A	1.00	0.00	0.00	0.80	0.40	1.20
2	B	2.67	0.33	0.67	2.40	1.20	2.93
3	C	3.33	0.67	0.33	3.20	1.60	3.47
4	D	2.00	0.00	0.00	1.40	1.20	1.60
5	E	2.00	0.00	0.00	2.00	1.00	2.00
6	F	2.00	0.00	0.00	1.60	0.80	1.40

表 5.8　胡贝尔点对之间影响指数

		1 A	2 B	3 C	4 D	5 E	6 F
1	A	2.00	0.00	0.00	0.80	0.40	1.20
2	B	2.67	1.33	0.67	2.40	1.20	2.93
3	C	3.33	0.67	1.33	3.20	1.60	3.47
4	D	2.00	0.00	0.00	2.40	1.20	1.60
5	E	2.00	0.00	0.00	2.00	2.00	2.00
6	F	2.00	0.00	0.00	1.60	0.80	2.40

由结果可见，这两个影响指数除了对角线不同之外，其余都相同。结果表明，对于行动者 A 来说，受到 C 的影响最大，B 次之。对于行动者 C 来说，受到 B 的影响最大。对于行动者 A 来说，他对 F 的影响最大。对于行动者 F 来说，他对 A 的影响也最大。

计算矩阵的行总和和列总和，便得到各个点的影响系数总和，如下所示：

表 5.9　卡兹各点影响指数

		1 Row S	2 Col S
1	A	3.40	13.00
2	B	10.20	1.00
3	C	12.60	1.00
4	D	6.20	11.40
5	E	7.00	6.20
6	F	5.80	12.60

表 5.10　胡贝尔各点影响指数

		1 Row S	2 Col S
1	A	4.40	14.00
2	B	11.20	2.00
3	C	13.60	2.00
4	D	7.20	12.40
5	E	8.00	7.20
6	F	6.80	13.60

可见,行动者 A 对所有其他人的影响指数总和为 3.40,而他受到的全部影响总和为 13,这表明,行动者 A 发出的影响远远没有得到的影响大。同理可见,对于行动者 C 来说,他发出的影响指数是 12.6,而得到的影响指数总和仅仅为 1,这表明,行动者 C“发号施令”的能力远远大于“遵守指令”的能力。

第五节　群体中心度

在一个大型组织中,中层管理人员是否比普通工人更居于组织网络的核心?在一个地区中,某个特定的少数族群是否比其他族群更容易形成一个社区?在中国的很多大学,如果某位校领导走马上任,他原先所在的研究团队或者他原来工作过的学院是否会随着他的“晋升”而在教学和科研方面得到越来越多的重视,该学院因而逐渐成为该大学的一个重要团队。如果这些成为“事实”,即说明一个大集体中的某个或某些小团体可能在整个集体中居于重要地位,即居于网络的核心,该小群体有其中心度。如何考察大群体中的一些小群体在多大程度上居于核心?如果将“中心度测度”应用到一系列行动者(而不是单个行动者),就可以解答这些问题。这需要我们将上文介绍的中心性指数推广到群体,这就是艾弗雷特等学者(Everett & Borgatti, 1999)的工作。

一、三类群体中心度指数的含义

(一) 群体度数中心度

对于一个大型网络中的某个小群体来讲,该群体的群体度数中心度(group degree centrality)指的是该小群体之外与该小群体内部成员有直接关联的点数。如果该群体有多个点与群体外的一点相连,则只计算一次。因此,在图 5.3 中,a 和 b 构成的群体的中心度是 6。

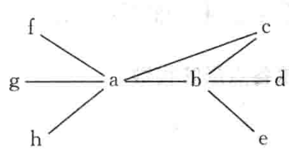


图 5.3　一个 8 点图

子群体的度数中心度的标准化方法是:将子群体的度数中心度除以非子本群体行动者的数目。因此,群体{a, b}的标准化点度中心度是 6/6 =1.0。

(二) 群体接近中心度

群体的接近中心度(group closeness centrality)可以界定为群体内部各点到群体外部各个点的距离之和。与个体接近中心度一样,这个测度也是一种反向

的测度，即接近中心度越大，说明该群体越不居于核心。

在具体计算的时候，涉及如何界定群体成员到外部成员之间的距离，这是可以讨论的。一个群体可以包含多个点，每一个点到群体外其他点都有距离，这些距离可以构成一个集合 D。我们可以分别把 D 中的最大值、最小值和平均值作为距离测度。

（三）群体中间中心度

群体中间中心度(group betweenness centrality)可以这样界定。假设 C 是一个图的子集，其中的点集记作 V。令 $g_{u,v}$代表连接点 u 和点 v 的捷径数，令 $g_{u,v}(C)$代表连接点 u 和点 v 的经过点集 C 的捷径数。那么，C 的群体中间中心度可记作 $C_B(C)$，其表达式为：

$$C_B(C) = \sum_{u<v} \frac{g_{uv}(C)}{g_{uv}} u, \; v \notin C$$

换句话说，群体中间中心度指的是经过该群体的连接非该群体成员之间的捷径数占捷径总数之比。为了测量这种测度，可采用如下方法：(1)计算任何一对非群体成员之间的捷径数，产生一个点—点数量矩阵；(2)删除所有涉及群体成员的关系，再重新计算，产生一个新的点—点矩阵；(3)新矩阵中的每一个值都除以第一个矩阵的相应值，得到各个比例值；(4)计算所有这些比例之和。

与个体的中间中心度类似，对群体中间中心度也可以进行标准化处理，方法是把每个值除以理论上的最大可能值。对于规模既定的群体来说，当它是星形结构的时候，最大可能值才出现。因此，我们定义标准化的群体中间中心度为 $C_B'(C)$。

$$C_B{}'(C) = \frac{2C_B(C)}{(|V|-|C|)(|V|-|C|-1)}$$

二、群体中心度指数的计算举例

下面用一个例子展示如何用 UCINET 计算群体中心度指数。该例是在 UCINET 软件中自带的数据，该数据是由林达·沃尔夫(Linda Wolfe)收集的，记载了佛罗里达州的 20 只猴子在三个月内的野外互动情况。具体的数据文件有两个，一个名为 WOLF，另一个名为 WOLFI。前一个数据文件中又包含两个数据矩阵 WOLFN 和 WOLFK。如果一些猴子共同出现在河边，便记录他们之

间有互动,这个数据就是 WOLFN。WOLFK 这个数据则记载了这些猴子之间假定存在的亲属关系,例如,18 号猴子可能是 19 号的孙女。WOLFI 这个数据是 20 行 4 列的数据,记载了 20 个猴子的四类信息:(1)身份代码,(2)年龄,(3)性别,(4)级别。

下面的表格是 20 个猴子的基本信息以及每个猴子的中心度分析的结果。

表 5.11 猴子个体的信息及其三种中心度

代码	性别	年龄	所在年龄群体	度数中心度	标准化的度数中心度	接近中心度	标准化的接近中心度	中间中心度	标准化的中间中心度
1	雄	15	14—16	4	21.05	142	13.38	1	0.58
2	雄	10	10—13	0	0	380	5	0	0
3	雄	10	10—13	13	68.42	133	14.29	44.5	26.02
4	雄	8	7—9	3	15.79	143	13.29	0	0
5	雄	7	7—9	2	10.53	144	13.19	0	0
6	雌	15	14—16	0	0	380	5	0	0
7	雌	5	4—6	3	15.79	143	13.29	0	0
8	雌	11	10—13	3	15.79	143	13.29	0.5	0.29
9	雌	8	7—9	1	5.26	145	13.1	0	0
10	雌	9	7—9	3	15.79	143	13.29	0	0
11	雌	16	14—16	2	10.53	144	13.19	0	0
12	雌	10	10—13	9	47.37	137	13.87	10.33	6.04
13	雌	14	14—16	6	31.58	140	13.57	1.83	1.07
14	雌	5	4—6	4	21.05	142	13.38	0	0
15	雌	7	7—9	6	31.58	140	13.57	1.83	1.07
16	雌	11	10—13	0	0	380	5	0	0
17	雌	7	7—9	3	15.79	143	13.29	0	0
18	雌	5	4—6	0	0	380	5	0	0
19	雌	15	14—16	0	0	380	5	0	0
20	雌	4	4—6	0	0	380	5	0	0

(一) 群体度数中心度

下面分别针对性别和年龄进行群体中心度分析。这里要用到六种不同的群体,前两个群体是基于性别形成的雌雄群体;后四个群体则是考虑到不同年龄的四个年龄段群体。由于初始数据是多值的对称数据,为了分析的方便,有必要先对数据进行二值化处理,二值化的规则是将互动次数多于 6 次的情况转变为 1,其余变成 0。不过,由于数据矩阵 WOLFN 包含在 WOLF 数据文件中,所以首

先要将 WOLFN 数据从 WOLF 数据文件中提取出来，这就要用到开包(unpack)命令，即在 UCINET 中，点击 Data→Unpack，选择 UCINET 安装目录下的 WOLF 文件，回车即可得到两个数据文件 WOLF-WOLFN 和 WOLF-WOLFK，前者就是我们要分析的数据 WOLFN。先对它进行二值化处理：在 UCINET 中，点击 Transform→Dichotomize，在出现的对话框中的 Input Dataset 项中选出 WOLF-WOLFN 文件，在 Dichotomization Rule 项目的 Value 上键入 6，回车即可得到二值化处理的数据，其命名为 WOLF-WOLFN_GT_6，此矩阵对应的网络图如下：

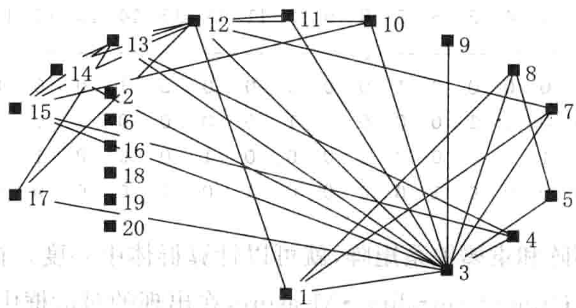


图 5.4　Wolf 猴群二值关系图

为了计算一个网络中各个子群体的中心度，需要有两个数据文件。一个文件是该网络中各个点之间的关系数据文件，另外一个便是记载了各个点隶属于哪个子群体的数据文件，该数据要指定哪些行动者隶属于哪个子群体。由于我们已经有了 WOLF-WOLFN_GT_6 这个数据文件，下面需要构造两个隶属关系数据矩阵，一个是性别矩阵，另外一个是年龄矩阵。

性别矩阵的构造方式如下。在 WOLFI 数据中可以看出，第 1，2，3，4，5 号猴子的性别是雄性，其余都是雌性。因此，要构造的性别矩阵就是一个 20 行 2 列的矩阵，20 行分别表示 20 只猴子，2 列表示性别，第一列表示雄性，第二列表示雌性。由于第 1，2，3，4，5 只猴子是雄性，因此，第一列的前 5 个数字都是 1，其余都是 0。第二列的前 5 个数字都是 0，其余都是 1。这样便构造了猴子—性别矩阵，该矩阵命名为 WOLF-WOLFN_GT_6 sex，其转置矩阵如下所示：

1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

年龄组矩阵的构造方式如下。在 WOLFI 数据中可以看出每个猴子的年龄(如表 5.11 所示)。要构造年龄群体矩阵,需要依据一定的标准。我们依据的标准是:4—6 岁为第 1 组,7—9 岁为第 2 组,10—13 岁为第 3 组,14—16 岁为第 4 组。据此构造的猴子—年龄组矩阵就有 20 行 4 列,第一列中有四个 1,表示第一组有四个成员(包括 7, 14, 18, 20),第二列有六个 1,表示第二组的成员有六人(包括 4, 5, 9, 10, 15, 17),其余列的含义依此类推。该矩阵命名为 WOLF-WOLFN_GT_6 age,其转置矩阵如下所示:

表 5.12 根据年龄分组的四组猴群

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
组 1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1
组 2	0	0	0	1	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0
组 3	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0
组 4	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0

有了数据矩阵和隶属关系矩阵,就可以计算群体中心度。在 UCINET 中,点击 Network→Group Centrality→Measure,在出现的对话框中的 files 框中的 Input Network 项上选出 WOLF-WOLFN_GT_6 数据文件,在 Input Groups:(actor by group indicator matrix)项目上选出 WOLF-WOLFN_GT_6 sex 文件,点击 OK 后,即可计算出结果如下:

表 5.13 雌雄猴群的群体中心度分析

Centrality Measures

	1 Degree	2 nDegree	3 2StepRe	4 n2StepR
1	10.00	0.67	10.00	0.67
2	4.00	0.80	4.00	0.80

由这个结果可见,由 5 个猴子构成的雄性群体的度数中心度为 10,标准化后的度数中心度为 0.67,由 15 个猴子构成的雌性群体的度数中心度为 4,标准化后的度数中心度为 0.80。在计算结果中还给出了群体 k-步可达(k-step reach)的值,本例中 k=2。所谓 k-步可达,等于从一个群体的成员出发,通过不超过 k 个步骤可达到的非该群体的成员数(或比例数)。从"2StepRe"可以看出,

第一个群体的成员通过不超过 2 步就可以达到另外一个群体的 10 个猴子。

如果在 Input Groups:(actor by group indicator matrix)项目上选出 WOLF-WOLFN_GT_6 age 文件,点击 OK 后,可计算出四个年龄组的中心度分析结果,如下所示:

表 5.14　年龄不同的四组猴群的群体中心度分析

```
Centrality Measures
             1          2          3          4
        Degree    nDegree    2StepRe    n2StepR
       -------    -------    -------    -------
1         5.00       0.31      12.00       0.75
2         5.00       0.36       8.00       0.57
3        11.00       0.73      11.00       0.73
4         8.00       0.53      11.00       0.73
```

对这个计算结果的解释与上述类似,此处不再赘述。上述两种情形计算的结果可以汇总如下。

表 5.15　六组猴群的群体度数中心度分析汇总

群　　体	成　　员	群体度数中心度	标准化的群体度数中心度
1. 14—16 岁年龄段	1　6　11　13　19	8	0.53
2. 10—13 岁年龄段	2　3　8　12　16	11	0.73
3. 7—9 岁年龄段	4　5　9　10　15　17	5	0.36
4. 4—6 岁年龄段	7　14　18　20	5	0.31
5. 雄性群体	1—5	10	0.67
6. 雌性群体	6—20	4	0.8

(二) 群体接近中心度

UCINET 仅仅计算出群体的度数中心度,并不提供计算接近中心度和中间中心度指标的指令。前文将群体接近度(group closeness)界定为从该群体出发向群体外成员发出的关系距离之和。从一点出发到多点的距离有多种,这些距离构成的集合可以命名为 D。我们可以界定一个点到一个点集合的距离为 D 中的最大值、最小值或平均值。例如,在图 5.3 中,如果采用最短距离法,那么由{8, 1, 7}构成的群体与点 12 的距离就是 1,因为点 1 与点 12 的距离就是 1。如

果采用最长距离法,那么该群体与点 12 的距离就是 2,因为点 7 和点 8 与点 12 的距离都是 1。当然,如果采用平均距离法,那么该群体与点 12 的距离就是 1.67,即(2+2+1)/3=1.67。当然,当群体仅仅由一个点构成时,所有上述距离测度都相等,群体中心度也就等于个体的接近中心度了。

群体接近中心度可以进行标准化,其处理方法是:接近中心度除以非本子群体成员的数目。标准化的结果是,大值表明大的接近中心度,表明该群体越在整个网络中占据中心地位。对于本例来说,六个群体的群体接近中心度及其标准化的值如下表所示:

表 5.16　六组猴群的群体接近中心度

群　　体	最小值	均值	最大值	标准化的最小值	标准化的均值	标准化的最大值
1. 14—16 岁年龄段	14	18	20	0.79	0.61	0.55
2. 10—13 岁年龄段	11	15	21	1	0.73	0.52
3. 7—9 岁年龄段	11	13.7	15	0.73	0.58	0.53
4. 4—6 岁年龄段	19	20.5	22	0.63	0.59	0.55
5. 雄性群体	10	16	20	1	0.63	0.5
6. 雌性群体	4	6.4	7	1	0.63	0.57

至于选择距离的最小值还是最大值作为接近中心度的指标,则取决于研究的问题。如果认为所研究的子群体是一个整体的单位(如公司、法人团体、交流网络等),不关注其内在的结构(特别是各个距离值),则可以利用最小值。相比之下,如果一个小群体的内部沟通不是特别好,或者内部沟通表现出一定的结构,则应该采用最大值。当子群体中的信息传递是从一个点出发随机向外传递的时候,可利用距离的均值作为群体接近中心度的指标。

(三) 群体中间中心度

下面分析这些群体的群体中间中心度。如前文所述,令 C 为一个图的子集,其点集记作 V。令 $g_{u,v}$代表连接点 u 和点 v 的捷径数,$g_{u,v}(C)$表示连接点 u 和点 v 的捷径中经过点集 C 的捷径数。那么,C 的群体中间中心度的表达式为:

$$C_B(C)=\sum_{u<v}\frac{g_{uv}(C)}{g_{uv}}\ u,\ v\notin C$$

也就是说,群体中间中心度指连接非该群体成员之间的捷径数中经过该群

体的捷径所占比例。前文介绍了测量这种测度的四种方法以及标准化处理方法。在本例中，按照这种方法计算得到的群体中间中心度及其标准化值如下表所示：

表 5.17　六组猴群的群体中间中心度

群　　体	群体中间中心度	标准化的群体中间中心度
1. 14—16 岁年龄段	2.84	0.03
2. 10—13 岁年龄段	43.50	0.41
3. 7—9 岁年龄段	0.00	0.00
4. 4—6 岁年龄段	0.00	0.00
5. 雄性群体	24.34	0.23
6. 雌性群体	0.50	0.05

由表 5.17 可见，两个群体(4—6 岁群体和 7—9 岁群体)的中间中心度都是 0，不过，只有 4—6 岁这个群体中的个体(包括 7，14，20)的中间中心度都是 0，而 7—9 岁群体(包括 4，5，9，10，15，17)中，第 15 号猴子的中间中心度并不是 0，而是 1.83。这意味着，本来该个体的中间中心度不是 1，但是加入该子群体之后，群体的中间中心度却变为 0，即某些个体可能“失掉”中心度。由分析结果还可看出，群体 2(2，3，8，12，16)和雄性群体(1，2，3，4，5)的中间中心度较高。在这六个子群体中，只有 6 个猴子(1，3，8，12，13 和 15)有非 0 的中间中心度，其中第 3 个猴子的中间中心度最大(见表 5.11)。通常来讲，如果一个子群体包含的个体具有较高的中心度，该群体会继承一部分较高的值。但是，如果大的中间中心度值是由于与群体外的成员之间的关系带来的，则不会这样。例如，第 3 个猴子拥有的大的中间中心度并没有使得其所在的雄性猴群的中间中心度变大。换句话说，有这样一种趋势，即一个子群体中的个体可以通过建立与子群体之外的个体之间的联系来提高自己的中间中心度。

本 章 小 结

在社会网络分析中，对权力的探讨集中体现在对“中心度”和“中心势”的量化分析上。本章介绍了个体的和群体的几类常用的中心度和中心势指数，包括：度数中心度、中间中心度、接近中心度，同时介绍了与它们相应的中心势指数。中心度刻画单个行动者在网络中所处的核心位置；中心势刻画的则是一个网络

所具有的中心趋势。对于一个拥有 n 个行动者的无向网络来说,中心度指数有 n 个(因为有 n 个点),但是计算出来的中心势指数只有一个。

“度数中心度”刻画的是行动者的局部中心指数,测量网络中行动者自身的交易能力,没有考虑到能否控制他人。“中间中心度”研究一个行动者在多大程度上居于其他两个行动者之间,因而是一种“控制能力”指数。“接近中心度”考虑的是行动者在多大程度上不受其他行动者的控制。如果网络中的一个行动者在交易的过程中较少依赖于他人,此人就具有较高的中心度。一个点越是与其他点接近,该点就越不依赖于他者。

第六章

凝聚子群分析*

社会学研究的一个重要任务是揭示社会的结构。尽管关于社会结构的研究有很多，例如，在社会理论、人类学、经济学等领域中都有相关论述，但是大多数研究都是理论上的定性说明，缺乏量化研究和操作化指标，而这正是社会网络分析的任务之一，社会网络领域主要是通过对群体的研究揭示社会结构的。

群体是社会心理学研究的重要内容。群体成员之间可以相互帮助、信息共享，在此意义上可以说群体比个体更有优势，不过要完成互助和信息共享，要求群体的成员认同群体，共同参与互动。社会心理学从群体与个体之间的互动角度出发，把个体看做群体中的行为者，他接受群体的观念、影响、规范和价值观等。社会心理学家贝尔斯(Robert Bales)于 20 世纪 50 年代最早开始了群体成员沟通的研究。他通过研究发现了三个重要现象：第一，群体中的讨论在对群体任务的讨论与群体成员之间的关系的讨论之间快速摇摆；第二，任务群体中的讨论也是变化的，从强调意见交换，到关照决策的内在价值，最后到进行决策本身；第三，不管群体的规模是多大，群体中最健谈者作出的评论数总是占评论总数的 40％至 50％，次健谈者作出的评论数占评论总数的 25％至 30％。其结果是，大的群体趋于被一两个人控制。

有关群体、子群、派系等概念及其研究发现在社会心理学和社会学中得到了广泛的应用。在社会心理学领域研究的群体一般是小群体(small group)，小群体一般指人数不多、相对稳定、有共同目的、相互接触较多的联合体，其规模一般为 3 至 20 人。而在小群体内部又可以分化出一些子群体(subgroup)，又叫做小圈子。小圈子的人数更少一些，“一般为 2 至 7 人”(时蓉华，2004：500)。尽管本书并不严格地将小群体的规模限定为 2 至 7 人，但是可以约定小圈子中的人数少于小群体中的人数。

小群体是社会学研究的重要内容之一。社会学从小群体在整个社会关系系

---

* 本章是在刘军(2004)第六章的基础上修订、增补而成的。

统中所占据的位置出发,分析小群体与大环境之间、各个小群体之间以及小群体内部各个成员之间的关系。在社会学领域中,群体的概念也多种多样。弗里曼(Freeman, 1992)曾回顾了群体概念在社会学领域中的应用,尽管如此,直到20世纪90年代,学者们并没有给出群体的精确的形式化定义,而这个工作的完成要归功于社会网络分析的学者。社会网络领域对群体的研究主要是从形式化的角度进行的,其目的是为了分析凝聚子群的各种类型,找到在一个整体网络中存在多少种凝聚子群。这种研究显然不同于社会心理学领域中的凝聚子群研究。

第一节 凝聚子群概说

广义地说,凝聚子群研究是一种社会结构的描述性研究。本节首先探讨社会结构的含义,重点指出社会网络意义下的社会结构的含义,然后给出分析凝聚子群的基础及凝聚子群的几种类型。

一、社会结构的质性分析和量化研究

对社会结构的研究存在两种研究思路:质性的(qualitative)结构观和量化的(quantitative)结构观。可以说,多数对社会结构的研究都属于"质的研究"。例如,有学者(周怡,2000)把关于社会结构的研究分为四个时期:社会学经典时期;帕森斯时代;后帕森斯时代以及后结构主义时代的社会结构研究。这些关于社会结构的概念和界定都属于思辨性的研究。"网络结构"范式则利用量化方法刻画社会结构的研究,也是"社会网络分析"学者所坚持的治学思路(Wellman & Berkowitz, 1988; Berkowitz, 1982)。

如何刻画一个群体的社会结构?齐美尔、霍曼斯、布劳等社会学家都进行过论述。本节介绍如何对群体这个概念进行操作化处理。由于多数网络研究者都坚持如下结构观,即社会结构是在社会行动者之间实存或者潜在的关系模式(Wellman & Berkowitz, 1988; Berkowitz, 1982; Scott, 2000)①。因此,结构

① 林南(2005/2001:33)从社会资本的角度给出社会结构的界定。他指出,社会结构由以下四个部分组成:(1)占据不同数量的一种或者多种有价值资源的一组社会单位(位置);(2)与权威(对资源的控制和获得)有等级性的关系;(3)在运用资源的时候共享一些规则和程序;(4)它们被寄托给遵循这些规则和程序的行动者。这种社会结构观把资源和权力结构结合在一起,而不仅仅是一种抽象的形式结构,因而具有实际意义。

分析者的一个重要关注点是分析出网络中存在的"子结构"(sub-structure)。网络分析者给出一系列算法,用来分析网络的整体结构如何由小群体结构(如n-派系、n-宗派、k-丛)组成。

群体是多种多样的,每个人的一生都生活在各类群体之中。尽管关于群体的定义有很多,我们给出在社会网络意义上的群体的定义,即群体是在即定目标和规范的约束之下,彼此互动、协同活动的一群社会行动者。

这个定义体现了群体概念的关键,即它的所有成员之间必须有互动。群体要有自己的目标,有群体规范、群体意识,还要有群体的分工。另外,群体一般要有一定的凝聚力,这体现在其成员要有认同感、归属感等(时蓉华,2004:491)。高度的凝聚力可以增强对群体目标的达成,促进成员遵守群体的规范,增强成员的自信心,也影响到群体的工作效率。另外,我们的群体定义不限于个人。或者说,群体中的"个体"可以是个人,也可以是家庭、小组、目标小群体、城市,甚至国家等。

在社会学研究特别是社会网络研究中,没有比较明确的"凝聚子群"的定义,它也不是一个有明确含义的概念。大体上说,"凝集子群是满足如下条件的一个行动者子集合,即在此集合中的行动者之间具有相对较强、直接、紧密、经常的或者积极的关系"(Wasserman & Faust, 1994:249)。正是由于我们可以从多个角度分析"较强、紧密、经常以及积极"等关系的属性,因此,凝聚子群的概念也有多种。本章介绍凝集子群的各种概念、基本特征及分析方法。

二、凝聚子群的类型

许多学者(Erickson, 1988; Friedkin, 1984;时蓉华,2004)指出,在社会学、社会心理学研究中,特别是在对有关社会现象(如社会团结、从众心理、精神健康、心理疾病等)的解释中,社会凝聚力(social cohesion)都扮演着重要角色。例如,福李金(Friedkin, 1984:236)认为,在社会学理论研究中,我们常常可以利用网络凝聚性(network cohesion)作为一个解释变量,特别是在研究群体的各个成员如何达成共识的时候,利用网络凝聚性作为一个解释变量更具有说服力。按照这种思想,我们可以期待在那些相互之间存在频繁接触的人们当中,同质性比较强;在那些相互之间交往比较少的人们之间,异质性就比较强。科林斯(Collins, 1988)也强调凝聚性在社会网络分析中的重要性:"个体与网络的关系越紧密,他们就越受到网络中的各种标准的影响。"孤立的和紧密关

联的群体构成了派系，在这种具有高度凝聚力的群体中，个体的信念常常很一致。

社会网络分析者的任务之一是对“社会群体”概念进行形式化处理。在他们看来，可以通过网络中行动者子集合的一些特征来刻画、研究社会群体。“子群”的形式化概念是通过子群体中成员的总体凝聚性给出的，这种凝聚性建立在成员之间的某些特定关系属性的基础上。然而，由于我们可以根据多种网络属性对一个群体的凝聚性进行量化处理，因此，凝聚子群也有多种形式化定义。在社会网络文献中，存在多种对凝聚子群进行量化处理的方法。总的来说，我们可以从四个方面考察凝聚子群，这四个方面也恰恰体现了网络分析者对凝聚子群进行形式化处理的四个角度(Wasserman & Faust，1994:251)：

(1) 关系的互惠性；

(2) 子群成员之间的接近性或者可达性；

(3) 子群内部成员之间关系的频次(点的度数)；

(4) 子群内部成员之间的关系密度相对于内、外部成员之间的关系的密度。

具体地说，建立在“关系的互惠性”基础上的凝聚子群考察的切入点是网络各个成员之间关系的相互性，也就是说，考察任何一对成员是否相互“选择”，是否为邻接点；建立在“成员之间的可达性”基础上的凝聚子群的关注点是子群的各个成员之间是否可达，但是不要求都邻接；建立在“成员之间关系的频次”基础上的凝聚子群分析的入手点是子群成员与其他成员之间的关系频次；建立在“相对频次”基础上的子群概念，则关注子群中的成员相对于网络中的其他行动者来说是否比较紧密，是否具有相对较高的凝聚力。

这几类凝聚子群概念的侧重点不同，与之对应的各种分析方法所关注的社会网络属性也不同。不过，这些概念测量的基础都是在群体中表现出来的关系的“模式”，而不是“内容”。换句话说，这里给出的四种界定的基础并不是现实世界中人们之间的实质意义上的凝聚性，而是在各种关系中表现出来的关系的数量特征，尽管实质意义上的凝聚性及其表现出来的数量特征是相关的(例如，在春节期间，子女通常到父母家里与亲人团聚，未必再短信拜年)。因此，如果研究人们在通过短信拜年中表现出来的凝聚力时，会发现亲子关系表现出来的“短信拜年”却很少，但这决不意味着亲子关系不具有凝聚力。亲子关系具有实质意义上的凝聚力，但未必表现出数量上的凝聚力。另外，同一个凝聚子群概念在不同性质的网络中的含义也不同。例如，“派系”这个概念在“无向关系网络”、“有向

关系网络”和“多值关系网”中的含义就有所不同。因此，下面将分别介绍不同情形下的凝聚子群含义。

需要指出的是，关于凝聚子群研究的文献更多地体现在社会心理学领域。例如，凝聚群体的目标、规范、群体心理、凝聚力、冲突、合作等内容已经得到相应的研究（时蓉华，2004），对这些问题感兴趣的读者请进一步参考有关资料，这里从略。

第二节 基于互惠性的凝聚子群

建立在互惠性基础上的凝聚子群主要是派系（cliques）。派系这个概念是在讨论图论的社会学意义时提出来的。最初，社会网络学者主要针对小群体中具有积极意义的关系（如朋友关系、喜爱关系等）进行凝聚子群分析，目的是找出可以“分派”的子群来。实际上，早期社会网络学者已经在霍桑工厂和杨基城报告中涉及“派系”思想。霍桑实验中涉及的“非正式群体”就可以看成是派系。一般来说，对于二元有向关系网络来说，“派系”常常指这样的一个子群体，即其成员之间的关系都是互惠的，并且不能向其中加入任何一个成员，否则将改变这个性质。派系是最基本的凝聚子群。对于不同类型的关系来说，利用图论语言可以给出派系的不同定义。

一、派系的定义

（一）无向关系网络中的派系

对于不同性质的网络来说，派系的定义也不同。在一个图中，“派系”指至少包含三个点的最大完备子图（maximal complete sub-graph）。这个定义有三点值得强调。首先，派系的成员至少包含三个点，之所以做如此要求，主要是为了强调一个互惠对（mutual dyad）不构成派系；其次，派系是“完备”的，即其中任何两点之间都直接相关，都是邻接的，并且不存在任何与派系中所有点都有关联的其他点；第三，派系是“最大”的，其含义是，我们不能向其中加入新的点，否则将改变“完备”这个性质。

“派系”是“最大的完全子图”，即在该点集中，任何点对之间都存在一条直接相连的线，并且该派系不能被其他任何派系所包含。如图 6.1 所示，一个包含 4 个成员的派系包含 6 条线，一个包含 5 个成员的派系有 10 条线，9 个成员的派

系有 36 条线。可以看出,一个派系中的任何两个成员之间都存在关系。

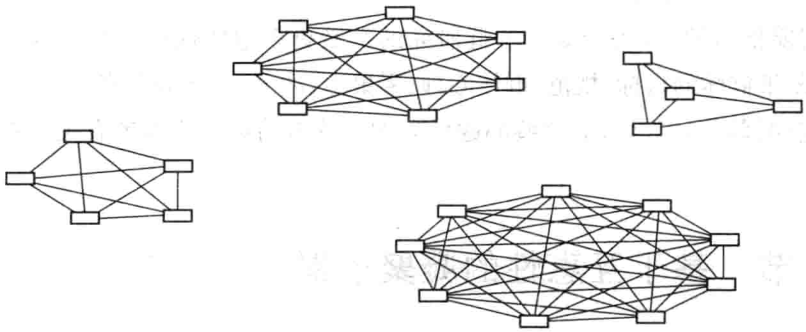


图 6.1　不同规模的派系

由上述分析可以总结派系拥有的如下性质:(1)派系的密度为 1;(2)一个包含 n 点的派系中任何一个成员都与其他 n－1 个成员相连;(3)派系中任何两点之间的距离都是 1;(4)派系内关系到派系外关系的比例达到最大;(5)派系中的所有三方组都是传递性三方组。

后文将指出,如果一个点集的任何两点都可以通过一定的途径相连,这样的点集叫做“成分”(component)。“派系”与“成分”不同,“派系”的概念要比“成分”严格。一个“成分”中的所有点之间不要求都是邻接的,而一个“派系”中的点必须都邻接。

(二) *有向关系网络中的派系*

上述派系概念只是针对无向网络图来说的。有向网络图中派系的含义稍有不同。在一个无向图中,所有点之间的关系都是相互的,考察派系要用到图中所有的线。在有向图中,根据派系的定义,只有“互惠”的关系才能纳入派系之中。因此,分析有向图中的派系一定针对的是行动者之间的互惠关系,且这种要求往往是比较严格的。

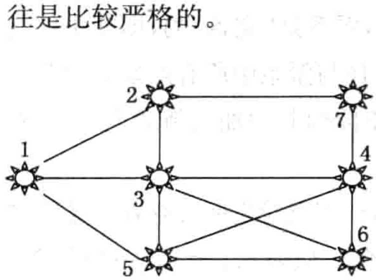


图 6.2　7 点关系图

例如,在图 6.2 中,{1, 2, 3}构成一个派系。假设该图是一个有向关系图,并且 1 和 2 之间的关系是相互的,2 和 3 之间的关系也是相互的,但是 1 和 3 之间只存在单向的关系,那么这三个点就不构成一个派系。可见,有向图中派系的概念是很严格的。因此,我们

称在有向图中区分出来的派系叫做强派系(strong cliques),把从无向图中分析得到的派系叫做弱派系(weak cliques)。

在针对具体数据分析其派系构成的时候,可以在 UCINET 中,点击 Network→Subgroups→Cliques,选择相关数据即可进行派系分析,找到其中有多少个派系以及每个派系包含哪些成员等。参见本章后文部分的分析。

(三) 多值关系网络中的派系

上文讨论的网络主要是二值关系网。但是,在现实生活中,关系常常是有多种取值的。例如,人们往往把朋友关系分为几类,如"密友"、"好友"、"一般朋友"和"一面之交的朋友"等。又如,在研究国际贸易的时候,贸易额绝不是"1"和"0"这样的二值数据,而是多值的。在一个多值网络中,凝聚子群研究的目的是为了找到其中的一些相互联系比较紧密、具有凝聚力的小群体。因此,子群成员之间的关系应该较多。

根据关系是否存在方向,可以把多值网络分为有向多值网和无向多值网两种。为了简单起见,下面对凝聚子群的分析主要针对的是无向多值网。

在研究无向多值网的凝聚子群(如派系、n-派系等)的时候,必须首先要确定该子群的凝聚程度,也就是要首先给出一个临界值 c 来。可以想象,该临界值越大(小),所发现的子群的凝聚力就越强(弱)。如果关系的取值在 0 和 C－1 之间的话,那么临界值的取值范围也在 0 和 C－1 之间。

1. "c 层派系"的含义。

既然关系是多值的,那么在多值网络中的派系概念就不只有一类,而是有多类。具体地说,在一个整体网中,一个 c 层派系(a clique at level c)指的是该图中的一个满足如下条件的子图,即该子图中任何一对点之间的关系强度都不小于 c,并且在子图外的任何一点到该子图中的所有点的关系强度都小于 c。

临界值 c 可以有不同取值,因此 c 层派系也可以有多种。需要注意的是,假设存在两个临界值分别为 c_1 和 c_2 的派系,并且假设 $c_1 > c_2$,那么 c_1 层次的派系将包含在 c_2 层次的派系之中(Doreian, 1969:240; Wasserman & Faust, 1994:279)。

2. "c 层派系"分析举例。

假设下面的数据代表某村落中 10 个家庭在一周时间内的相互拜访关系(本

表参考了 Doreian，1969:241)。下面我们分析其中存在的多值派系的情况。

表 6.1　10 个家庭在一周时间内的互访关系

	A	B	C	D	E	F	G	H	I	J
A	0	5	0	1	0	1	5	4	1	1
B	1	0	3	5	1	5	3	1	3	5
C	0	4	0	5	1	5	1	0	0	0
D	5	5	3	0	3	5	5	0	3	5
E	0	5	1	1	0	1	3	0	4	1
F	3	5	3	5	5	0	5	4	3	5
G	4	5	0	3	1	5	0	4	3	4
H	5	0	1	1	0	0	5	0	5	3
I	4	4	1	3	1	5	5	0	0	2
J	1	5	0	4	3	5	4	1	2	0

可以利用一定的计算程序计算出该网络包含的各种派系的情况。当 c＝1 的时候，即在第一个层次的意义上，可以发现存在四个 1 层次派系：{B, C, D, E, F}、{A, G, H, J}、{A, B, D, F, G, I, J}和{E, B, D, F, G, I, J}。

当 c＝2 的时候，可以发现存在的 2 层次派系要比第一层次的派系稍有改变，它包括：{A, G, H}、{B, C, D, F}和{B, D, F, G, I, J}。

当 c＝3 的时候，可以发现存在的 3 层次派系包括：{A, G, H}、{B, C, D, F}，{B, D, F, G, I}和{B, D, F, G, J}。

当 c＝4 的时候，在上一步分析中发现的包含 5 个成员的 3 层次派系{B, D, F, G, I}和{B, D, F, G, J}则分裂为两个分别包含 4 个成员的 4 层次派系，变为{B, F, G, I}和{B, D, F, J}。

当 c＝5 的时候，层次已经达到最高，5 层次派系只包括{B, D, F}、{B, F, G}和{B, F, J}。可见，点 B 和点 F 都存在于上述三个派系之中。

由上述这种派系分析，我们可以清楚地看到派系之间相互包含的情况。上述分析可总结为表 6.2 所示。

表 6.2　村民之间互访关系的派系结构

派系的层次	每个层次的派系
1	{A, B, D, F, G, I, J}, {B, E, D, F, G, I, J}, {B, C, D, E, F}, {A, G, H, J}
2	{B, D, F, G, I, J}, {B, C, D, F}, {A, G, H}
3	{B, D, F, G, J}, {B, D, F, G, I}, {B, C, D, F}, {A, G, H}
4	{B, D, F, J}, {B, F, G, I}
5	{B, D, F}, {B, F, G}, {B, F, J}

如果说规模较小的网络比较容易找到派系的话，那么在网络规模变大的情况下，一般难以分析其中存在的派系。这时候研究者只能通过电脑编程，利用计算机找到各种层次的派系。在 UCINET 中，不能直接分析多值关系网络中的派系，但是可以间接地计算出来。例如，可首先对数据分别根据不同的临界值（如1，2，3等）进行二值化处理，得到多种二值数据矩阵。然后分别对这些矩阵进行派系分析，即可得到当c取不同值时的c层次派系结果。

二、对派系概念的评价

如果研究网络中关系相当紧密的小群体的话，“派系”无疑是可以利用的一种凝聚子群概念。但是，学者们常常批评派系概念具有如下相互关联的四个缺点。

首先，派系这个概念太严格。只要去掉一个关系，派系就不成其为派系了。另外，在有向关系网和多值关系网络中，c层次派系的含义也很严格。对于一个有向关系网中的一个小群体来说，只有当它的成员之间的关系都是互惠关系时，该群体才可能成为一个派系。在多值关系网络中，只有其成员之间的关系取值都不小于c的时候，才可能成为c层次的派系。显然，这种定义很严格，在现实生活中也比较少见。

其次，派系的规模受到点的度数的限制。如果行动者被限定为拥有k个关系，那么，在结果中就不可能包含超过k个成员的派系。

再次，现实中出现的派系常常规模太小，并且重叠很多。在一个联系比较稀疏的网络中，派系的数目可能极少。特别是在有向关系网络中，派系的数目更少。

最后,派系概念的另一个缺陷是,派系的成员之间没有任何分化(Doreian, 1969; Freeman, 1992)。也就是说,既然派系是最大的完备子群,那么派系中的所有成员在图论的意义上都是等同的,所有成员之间都邻接。假如我们期望一个网络中的凝聚子群表现出一些令人感兴趣的结构特点来(如群体内存在得到广泛认同的核心成员,存在一些不太被认同的边缘成员等)。这时候,派系的概念就不是一个恰当的凝聚子群概念。对于一个多值关系网络来说,其中一个 c 层次派系中的各个成员之间的关系都不小于 c。换言之,c 派系的概念把所有这些大于 c 的关系看成是等同的,并不关注它们之间的进一步差异,不关注它们之间的进一步分化。这是派系概念的缺点之一。

另一方面,一些研究大网络(如包含几百个行动者的网络)的学者们也发现,群体中可能存在大量相互重叠的派系。在这种情况下,派系的概念就没有多少意义了。这时候应该研究派系的重叠性。

正是由于派系这个概念太严格,学者们在此基础上给出了另外一些凝聚子群概念,使之具有较大的理论价值和应用价值,这就引出了一系列推广的凝聚子群概念。

总的来说,学者们利用两种结构特征放松了派系概念的严格性。首先,艾尔巴(Alba, 1973)等利用"可达性"、"路径距离"和"直径"这几个性质推广派系的概念;其次,有学者利用"点的度数"(nodal degree)提出另外一些凝聚子群概念。这两种推广的起点都是派系,都是通过放松一个或者几个限制性因素来推广派系概念的(参见 Scott, 2000:100—122; Wasserman & Faust, 1994:253—256; Everett, 2002)。这一点将在下一节进行论述。

第三节 基于可达性的凝聚子群

在现实生活中,行动者之间的关系未必是直接的,也可能是间接的。例如,谣言的传播,通过关系找工作等社会过程往往需要第三者参与,而这些行动者之间的关系往往不是很紧密。因此,有理由相信,联系不是很紧密的一个小群体也可以叫做"凝聚子群"。这时候,建立在可达性(reachability)等基础上的两个凝聚子群概念,即 n-派系和 n-宗派就可派上用场。正是考虑到这一点,以下介绍这两种凝聚子群。

一、n-派系

如前所述，派系是最大的完备子图，这个概念比较严格。因此，有人从"可达性"这个角度对此概念进行了推广。建立在可达性基础上的凝聚子群考虑的是点与点之间的距离，要求一个子群的成员之间的距离不能太大。这样，我们可以设定一个临界值 n 作为凝聚子群成员之间距离的最大值，这就引出了对派系概念做出最早推广的 n-派系(n-cliques)概念。与上述类似，无向网络和有向网络中的 n-派系的含义有所不同。下面分别进行研究。

(一) 无向网络中的 n-派系

1. 无向二值网络中 n-派系的定义。

对于一个无向二值关系网络图来说，如果其中的一个子图满足如下条件，就称之为 n-派系：在该子图中，任何两点之间在总图中的距离(即捷径距离)最大不超过 n。从形式化角度说，令 d(i, j)代表两点 n_i、n_j 在总图中的距离。那么，一个 n-派系的形式化定义就是一个满足如下条件的拥有点集 Ns 的子图，即：$d(i, j) \leqslant n$，对于所有的 n_i，$n_j \in Ns$ 来说，在总图中不存在与子图中的任何点的距离不超过 n 的点。

n-派系的概念比较接近人们日常对派系的理解。这里的 n 是派系成员之间距离的最大值。如此看来，一个 1-派系实际上就是最大的完备子图本身，即上述的"派系"。一个 2-派系则是这样的一个派系，即其成员或者直接相连(距离为 1)，或者通过一个共同邻点间接相连(距离为 2)。n＝2 也是一个较好的临界值。当然，在实际的 n-派系分析中，n 的大小要由研究者自己决定。n 越大，对派系成员限制的标准就越松散。例如，一个 3-派系就是一个比 2-派系松散的群体。n 的最大值不能超过图中点总数减去 1。

n-派系可以通过矩阵操作方法进行分析，这在许多社会网络分析软件中都存在。例如，仅仅对邻接矩阵进行操作，就可以产生一个距离矩阵。这个矩阵的布尔代数平方表示的是所有距离为 2 的关系，三次方给出所有距离为 3 的关系，依此类推。大多数社会网络分析软件都可以进行这种计算。

我们也可以分析多值图中的派系，方法与前一部分所讨论的一样，用一种临界点标准进行分析。对于 n 的每一个取值来说，这种分析都会产生一系列嵌套派系：嵌套 2-派系，嵌套 3-派系等(可参见 Scott，2000；Wasserman & Faust，1994)。有关无向网络中的 n-派系分析举例，请读者参见刘军(2004:161)。

（二）有向网络中 n-派系

在介绍有向网络中的 n-派系之前，有必要先介绍一些新的概念，这就是有向关系图中的“半途径”(semipath)和“关联度”(connectivity)。这两个概念分别是无向网络中的“途径”和“关联度”概念的推广。

在一个有向关系网络中，一条点 i 到点 j 的“途径”指的是从 i 点出发指向 j 点、方向相同，点和线却不同的点—线系列。“途径”的长度指其中包含的线的条数。点 i 到点 j 的一条“半途径”(semipath)指的是从 i 点出发指向 j 点，由各不相同的点和线构成的系列。也就是说，有向网络中的途径要考虑到关系的方向，而“半途径”则不考虑方向。半途径的长度也是其中包含的线数。

1. 有向网络中的四类关联性。

这样看来，对于一对点 i 和点 j 来说，连接二者之间的长度为 n 的途径可能有如下四类性质(Wasserman & Faust, 1994:275—276)。

(1) n-弱关联性(weakly n-connected)，如果它们之间通过一条长度不超过 n 的半途径连接在一起；

(2) 单向 n-关联性(unilaterally n-connected)，如果它们之间存在一条从 i 指向 j，并且长度不超过 n 的途径，或者存在一条从 j 指向 i，并且长度不超过 n 的途径；

(3) n-强关联性(strongly n-connected)，如果它们之间存在一条从 i 指向 j，长度不超过 n 的途径，并且存在一条从 j 指向 i，长度不超过 n 的途径。从 i 指向 j 的途径中包含的点和线可能不同于从 j 指向 i 的途径中的点和线；

(4) n-递归关联性(recursively n-connected)，如果它们是 n-强关联的，并且从 i 指向 j 的途径与从 j 指向 i 的途径包含相同的点和线，只是方向相反。

这四种关联的含义依次递进，越来越严格。对应上述四类关联性，有向关系网络中的 n-派系也有如下四类。

2. 有向网络中的四类 n-派系。

一个子图可能是如下四类 n-派系之一，只要它满足相应的条件。

(1) 弱关联的 n-派系(a weakly connected n-clique)，如果该子图中的所有点都是 n-弱关联的，并且不存在与该子图所有点之间具有 n-弱关联关系的其他点；

(2) 单向关联的 n-派系，如果该子图中的所有点都是单向 n-关联的，并且不存在与该子图所有点之间具有单向 n-关联的其他点；

（3）强关联的 n-派系，如果该子图中的所有点都是 n-强关联的，并且不存在与该子图所有点之间具有单向 n-强关联的其他点；

（4）递归关联的 n-派系，如果该子图中的所有点都是递归 n-关联的，并且不存在与该子图所有点之间也具有递归 n-关联的其他点。

（三）多值网络中的 n-派系

在一个多值关系网中，n-派系建立的基础是子群成员之间的各个捷径之值。上一节介绍了多值关系网络中的 c 层次途径的含义。与之类似，在一个多值关系网中，如果两点之间存在一条 c 层次途径的话，我们就说这两个点是在 c 层可达的（reachable at level c）。c 层次的 n-派系就是如下这种子群，即其中所有点之间的捷径上的所有线的取值都不小于 c。因此，在一个多值图中，一个 c 层 n-派系就包含了所有那些相互之间通过长度不超过 n 的途径在 c 层次上可达的成员，或者说其中的每对成员之间都存在一条 c 层次的派系，即其中所有点之间的捷径上的所有线的取值都不小于 c。

有向网络与多值网络中的 n-派系分析思路相同。只要遵循定义，有向网络以及多值网络中的 n-派系都是可以找到的，具体例示从略。

（四）对 n-派系的评价

n-派系思想在应用方面有三个局限性。首先，当 n 大于 2 的时候，很难给它以社会学的解释。其次，作为一个子图，n-派系的直径有可能大于 n。因为，n-派系虽然要求各个点之间通过长度不超过 n 的路径连接在一起，但这并不能保证这些路径仍然保留在子群之中。再次，一个 n-派系可能是一个不关联图（参见刘军，2004：162）。因此，对 n-派系概念进行推广便在情理之中，n-宗派就是一个推广的概念。

二、n-宗派

艾尔巴（Alba，1973）和莫坎（R.J.Mokken）等学者对 n-派系的思想进行了推广。莫坎提出 n-宗派（n-clan）这个概念。他指出，所谓 n-宗派指的是满足如下条件的一个 n-派系，即其中任何两点之间的捷径距离都不超过 n。可见，所有的 n-宗派都是 n-派系。具体例子参见刘军（2004：图 6.5）。

n-宗派与 n-派系都是子图，二者之间最主要的区别在于对“距离”的理解上。后者所说的“距离”指的是两点在“总图”中的距离，前者指的是两点在“子图”中的距离。应该说，n-宗派概念比 n-派系概念严格一些。或者可以说，任何 n-宗派

都是 n-派系,反之不成立。

第四节　基于度数的凝聚子群

除了上述凝聚子群之外,还有另外一些凝聚子群。本节介绍建立在点度数基础上的凝聚子群,这种凝聚子群是通过限制子群中的每个成员的邻点个数而得到的。

一、k-丛

与派系概念不同,以度数为基础的凝聚子群研究要求其成员的度数满足一定的条件,这就是赛德曼和弗斯特(Seidman & Foster, 1978)提出来的 k-丛(k-plex)概念。

赛德曼和弗斯特观察到,n-派系常常不稳健,表现出“脆弱性”(vulnerability)。所谓一个图是稳健的,其含义是,去掉该图的一个或者几个点之后,图的结构不受到太大的影响,否则称该图是不稳健的。我们可以利用关联性(connectivity)来测量稳健性,即对于一个图来说,“如果移去任何一个点,那么该图的结构在多大程度上是脆弱的”。在图 6.3 中,每一条径都包含点 3,点 3 对于任何其他点对之间的联系来说都至关重要。如果没有点 3,那么点 1、点 2、点 4 相互之间都不能联系在一起。因此,这个 2-派系是“脆弱的”。

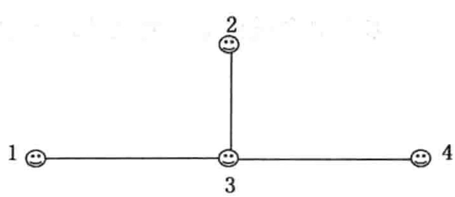


图 6.3　一个不稳健的 2-派系

(一) 无向网络中的 k-丛

通过这种分析,赛德曼和弗斯特提出了如下 k-丛概念:一个 k-丛就是满足下列条件的一个凝聚子群,即在这样的一个子群中,每个点都至少与除了 k 个点之外的其他点直接相连。也就是说,如果一个凝聚子群的规模为 n,那么只有当该子群中的任何点的度数都不小于(n－k)这个值的时候,我们才称之为 k-丛。

如果 k＝1，根据定义，1-丛中的每一个成员都与其他 n－1 个点相连，那么，一个1-丛就是一个派系，是一个最大的完备子图。当 k＝2 的时候，其中所有点都至少与 n－2 个其他点相连，但是，2-丛可以不是 2-派系。在图 6.4 中，图(1)是一个 3-派系，因为所有点之间的距离都不大于 3。然而，它却不是一个 3-丛，因为与点 A、C、E、F 相连的成员的数目都少于 6－3＝3。图(2)则既是一个 3-派系，也是一个 3-丛。

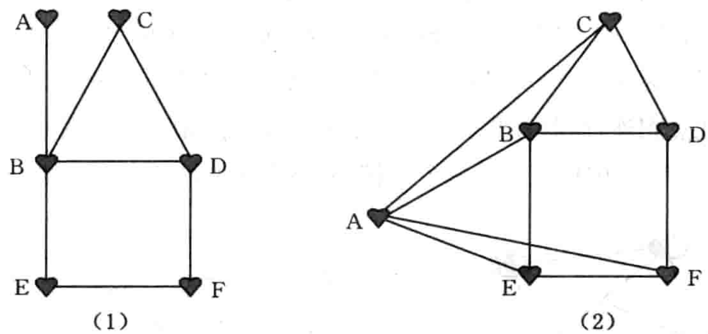


图 6.4　一个 3-派系和 3-丛

在分析 k-丛的时候，一个重要的问题是，研究者如何确定 k-丛的最小规模，即 k 是多大的时候才可以接受。当 k 取值较小的时候，k-丛可以是相对较小的。但是，k 值的增大会产生无意义的结果。其原因是，k 值高的子图都只能是内聚力较小的子图。那么，一个可以接受的 k-丛的最小规模应该是多少？表 6.3 给出了在这个问题上学者们坚持的经验"原则"(Everett，2002)。

表 6.3　可接受的 k-丛的最低规模

k	k-丛的最低规模
2	4
3	5
4	7
k	2k－1

（二）多值网络中的 k-丛

在多值网络中，k-丛的定义稍微复杂一些，需要考虑 c 层次的 k-丛这个概念。具体地说，对于多值关系网络中的一个凝聚子群来说，如果其中的全部 g_s 个点到该子群的至少 $g_s－k$ 个点之间的关系的取值都不小于 c 的话，这种凝聚

子群就叫做 c 层次的 k-丛(Wasserman & Faust，1994:279)。

多值网络中的 k-丛的具体分析过程从略。只要遵循 k-丛的定义，k-丛就是可以找到的。k-丛这个概念比 n-派系更能体现凝聚力思想，当 n 的取值比 2 大的时候更是如此。

二、K-核

与 k-丛概念相对的另一个建立在点度数基础上的凝聚子群概念是 k-核(k-core)。如果一个子图中的全部点都至少与该子图中的其他 k 个点邻接，则称这样的子图为 k-核。k-核与 k-丛不同，后者要求各个点都至少与除了 k 个点之外的其他点相连，而前者要求任何点至少与 k 个点相连。在图 6.4 中，②也是一个 3-核。图 6.5 中包含左右两个 3-核，而整个图也是更大图中的一个 3-核。

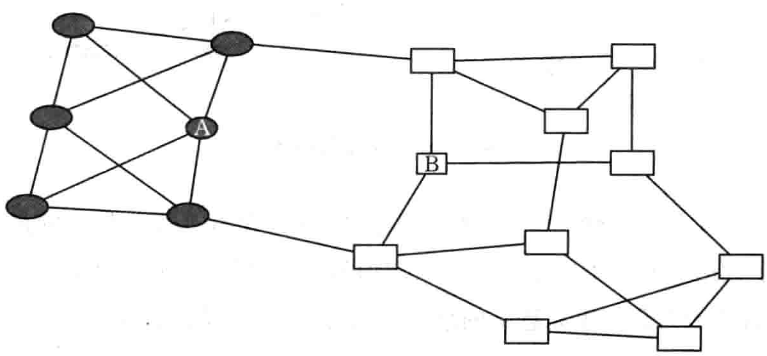


图 6.5　两个 3-核

一般来说，作为一类凝聚子群，k-核有自己的优势所在。k 值不同，得到的 k-核也不同。研究者要根据自己的数据和研究目的自行决定 k 值的大小，可以从中发现一些有意义的凝聚子群，这是研究 k-核的一个好处。k-核不一定是具有高度凝聚力的子群，但它们能够表现出与派系类似的性质来。

在 UCINET 中，点击 Network→Regions→K-Core 后出现对话框，选出要分析的数据(如“帮工关系第一子群.##h”数据)，点击 OK 后会计算出如下结果。

可见，在全部 25 个村民家庭之间的帮工关系中，可以进行四种分区，其度数分别为 4，8，9，10。对于度数为 4 的分区(即 4-核)来说，它包含 24 个村民家庭，只有第 13 号家庭被排除在外，因为该家庭仅给其他家庭提供 2 次帮工，不能纳入 4-核的定义之中。这 24 个家庭的度数至少是 4。度数为 10 的核(即 10-核)

表 6.4　k-核分析结果

```
K-CORE
--------------------------------------------------------------------

Input dataset:                 H:\whole UCINET\数据\帮工第一子群
Output partitions:             KCores

Warning: Matrix symmetrized by setting Xij and Xji to Max(Xij,Xji)

K-CORE PARTITIONING

                        1 1 1   1 1 1 2 2 2 2 1 2 2 1 1 1
         1 2 9 4 5 6 7 8 2 1 7 3 0 8 9 2 3 4 5 6 0 1 4 5 3

                        1 1 1   1 1 1 2 2 2 2 1 2 2 1 1 1
Degree   1 2 9 4 5 6 7 8 2 1 7 3 0 8 9 2 3 4 5 6 0 1 4 5 3
------   - - - - - - - - - - - - - - - - - - - - - - - - -
    10   XXXXXXXXXXXXXXXXXXXXX . . . . . . . . . . . . . .
     9   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX . . . . . .
     8   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX . . .
     4   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX .
```

中包含的家庭有 11 家,分别是 1, 2, 4, 5, 6, 7, 8, 9, 11, 12, 17。换句话说,这 11 个家庭之间都是两两相连的。

表 6.5　k-核分析的分区结果

```
Partition Metrics - each column is a partition

                      1       2       3       4
                 ------  ------  ------  ------
 1 nClusters      2.000   4.000   7.000  15.000
 2       CL1      0.960   0.880   0.760   0.440
 3       CL2      0.040   0.040   0.040   0.040
 4       CL3              0.040   0.040   0.040
 5       CL4              0.040   0.040   0.040
 6       CL5                      0.040   0.040
 7       CL6                      0.040   0.040
 8       CL7                      0.040   0.040
 9       CL8                              0.040
10       CL9                              0.040
11      CL10                              0.040
12      CL11                              0.040
13      CL12                              0.040
14      CL13                              0.040
15      CL14                              0.040
16      CL15                              0.040

LEGEND:
  (1)nClusters = number of clusters in the partition
  (2)CL1 = proportion of items that fall in cluster 1
  ...
  (n)CLn = proportion of items that fall in cluster n
```

上面这个表给出了在上述四种 k-核分区中包含的聚类数。可见,这四种分区中的聚类数分别是 2, 4, 7, 15。例如对于第一列来说,它包含两个聚类,实际上就是第 13 号家庭作为一个聚类和另外一个聚类,即 4-核。对于第四列来说,它包含 15 个聚类,聚类 1(CL1)由上述 11 个家庭构成,另外 14 个聚类分别由剩下的

14 个家庭构成。

表 6.6 给出的是每个点的核心度(KCores)以及每个点的四种分区情况。

表 6.6 每个点的核心度和分区情况

Coreness values for each node

	Page
1	10.00
2	10.00
3	9.00
4	10.00
5	10.00
6	10.00
7	10.00
8	10.00
9	10.00
10	9.00
11	10.00
12	10.00
13	2.00
14	4.00
15	4.00
16	8.00
17	10.00
18	9.00
19	9.00
20	8.00
21	8.00
22	9.00
23	9.00
24	9.00
25	9.00

Input dataset: H:\whole UCINET\数据\KCores

	1	2	3	4
1	1	1	1	1
2	1	1	1	1
3	1	1	1	2
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	3
11	1	1	1	1
12	1	1	1	1
13	2	2	2	4
14	1	3	3	5
15	1	4	4	6
16	1	1	5	7
17	1	1	1	1
18	1	1	1	8
19	1	1	1	9
20	1	1	6	10
21	1	1	7	11
22	1	1	1	12
23	1	1	1	13
24	1	1	1	14
25	1	1	1	15

在计算结果中,有一个输出的文件包含一个行动者 k-核分区矩阵(k-core by actor partition matrix)KCores,其定义如下:第 i 列第 j 行中的 k 值意味着点 i 就 j-核分区而言处于分区 k 中。例如,第 5 行第 4 列的值是 1,这说明第 5 个点位于第一种分区之中。

在每个点的核心度指标(Coreness values for each node)中,只有一列,它记载了每个点位于不同的 k-核的情况。例如,第 5 行的值是 10,这说明点 5 位于 10-核之中;第 13 行的值是 2,这说明点 13 位于 2-核之中。

第五节 基于"子群内外关系"的凝聚子群

艾尔巴(Alba, 1973)指出,一个凝聚子群至少涉及两个方面:一是关注子群内部的关系;二是比较子群内部成员之间的关系频次相对于子群内、外部成员之间的关系频次的大小。艾尔巴把子群内部的关系与子群之间的关系称为凝聚子

群的“核心—边缘”(centripetal-centrifugal)维度。在这种思想引导下,一些学者开始比较子群内部关系的紧密性与子群外部关系的稀疏性,这就引出下面将要介绍的几个凝聚子群概念。

一、成分

如果一个图可以分为几个部分,每个部分的内部成员之间存在关联,而各个部分之间没有任何关联,在这种情况下,我们把这些部分称为成分(component)。在一个图中,如果拿走其中的某个点,整个图的结构就分为两个互不关联的子图(成分)的话,那么则称该点为切点(或桥点)(cutpoint)。可以想象,切点在网络中占据重要的位置,对于其他点之间的沟通和信息传递来说具有重要意义。该点所代表的行动者往往是重要的行动者,他扮演着“掮客”或者“中间人”的角色,起到“中介”的作用。有关中间人的分析请参见本书第九章。

在 UCINET,点击 Network→Regions→Components→Binary Graph(或 Simple graphs、Valued Graphs)就可以进行成分分析。该程序可以分析二值网络数据、多值数据以及无向关系网、有向关系网络中的成分。需要注意的是,在有向网络数据中,成分的定义稍微复杂一点,分为“弱成分”和“强成分”两种。简言之,在有向网络中,如果忽略关系的方向,分析得到的成分就是“弱成分”;如果在一个成分中,任何两点之间都存在严格双向的途径,这样的成分就叫做“强成分”。

可以将一个孤立点看成是一个成分,整个网络也可以看成是一个成分。我们一般对既非孤立点,也非整体网的成分更感兴趣,因为这样的成分反映了整个网络可以分为多少个小群体。当然,如果不对“成分”这个概念进行一些限定的话,它也就没有太多的理论意义和现实意义。其原因在于,该定义没有给出这个独立存在的小群体到底是怎样的,因为我们是根据一个小群体是否独立存在来判断它是否为一个成分的。成分的定义不能告诉我们其他信息,例如该成分的内部成员之间的联系是多还是少?内部是否有派别?成分的成员之间平均距离是多少?因此,需要对“成分”这个凝聚子群概念进行推广。“块”就是这样的一个推广的概念。

二、块

如果一个图分为一些相对独立的子图的话,则称各个子图为“块”(blocks)。

用来构建“块”的程序模型叫做块模型(blockmodels)。

块模型最早是由怀特、布尔曼和布雷格(White, Boorman & Breiger, 1976)提出来的,它是一种研究网络位置模型的方法,是对社会角色的描述性代数分析。块模型是在位置层次上的研究,而非个体层次上的研究(参见 Wasserman & Faust, 1994:394—396; Burt, 1976:93—122)。有关块模型的定义和分析方法,请参见本书第八章,此处不再赘述。

三、LS 集合

如果既考虑子群内部关系的频次,也考虑子群内成员向子群外发出关系的频次,就可以得到一个新的子群概念,这就是 LS 集合。

(一) LS 集合的含义

赛德曼(Seidman)(转引自 Wasserman & Faust, 1994:268)是这样界定 LS 集合的:

> “在一个社会网络中,如果存在满足如下条件的一个点集 S,则称该点集为 LS 集合:如果该点集内的每个真子集合中存在的到‘该真子集合在 S 中的补集’的关系都多于该真子集合到‘S 外’的关系。”

下面用形式化的语言表述 LS 集合的含义。假设存在一个图 G,它对应的点集为 N; G 中存在一个子图 G_s,其对应的点集记作 N_s;再假设 N_s 中存在一个真子集合 S_s,即 $S_s \subset N_s$。如果任何一个真子集合 S_s 到 $N_s - S_s$(即 S_s 在 N_s 中的补集)中点的关系都多于到 $N - N_s$ 中点的关系的话,则称 N_s 是一个 LS 集合。

可见,LS 集合的建构要比较子集合内部和子集合之间的关系。这里要考虑的基本集合有三个,它们满足如下关系:$S_s \subset N_s \subset N$。N 是总的点集合,即对应于总图 G 的点集合。首先需要注意的是,子集合 S_s 代表的是从 N_s 中选出来的任何一个子集合。N_s 是一个潜在的 LS 集合。其次要考虑两个补集合 $N - N_s$ 和 $N_s - S_s$。还要考虑两类线:从 S_s 到集合 $N_s - S_s$ 的线和从 S_s 到集合 $N - N_s$ 的线。LS 集合 N_s 内部的线应该比从 LS 集合中的一个真子集合到非 LS 集合的成员的线要多得多。

LS 集合拥有如下重要性质。首先,由于 LS 集合中的所有子集合内部的关系都要多于外部的关系,因此它们都是相对稳健的(robust),不包含分裂性的群体。因此,我们可以假设一个社会网络中的 LS 集合会随着时间的推移保持相对的稳定。其次,在一个既定的图中可能存在多个 LS 集合,在各个 LS 集合之

间有一种重要的关系，即任何两个 LS 集合或者没有任何共同的成员，或者一个 LS 集合包含另外一个 LS 集合。可见，这一点与派系、n-派系和 k-丛等概念不同，因为这些子群是可以相互交叉的，但是，LS 集合不能交叉重叠。LS 集合之间或者是包含与被包含的关系，或者无公共点，这个事实意味着在一个图中存在着一系列等级嵌套的 LS 集合。

（二）LS 集合分析举例

LS 集合这个概念的本质含义可以看成是其子集合的合集。最初，学者们利用晶体管中的芯片来展示 LS 集合的构成。这种晶体管设计的目的是在各个芯片之间建立最小数目的关联。因此，LS 集合这个概念保证了如果在一个 LS 集合中包含的任何一个子集合被安装在芯片上，它与外界建立的关联就要多于 LS 集合与外界建立的关联。

因此，集合 N_s 中的所有点就首先构成一个 LS 集合。N_s 中的每个单独的元素 n 也可以看成是一个 LS 集合，并且可以称之为"平凡 LS 集合"（trivial LS sets）。

上文已经指出，LS 集合中所有子集合内部的关系都要多于外部的关系，因此它们都相对稳健，一般不会分裂。例如，在图 6.6 中包含 16 个 LS 集合（引自 Borgatti，Everett & Shirey，1990：340）。

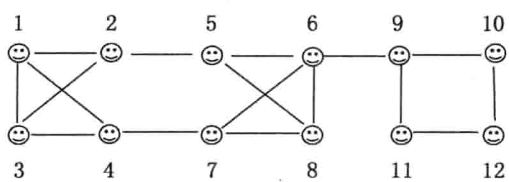


图 6.6　一个 12 点图

这 16 个 LS 集合分别是：

{1}，{2}，{3}，{4}，{5}，{6}，{7}，{8}，{9}，{10}，{11}，{12}，{1 2 3 4}，{1 2 3 4 5 6 7 8}，{9 10 11 12}，{1 2 3 4 5 6 7 8 9 10 11 12}

可见，除了单点集合之外，其余四个集合中的个体成员相互之间的联系都比与外部成员之间的联系多。这一点恰恰是有凝聚力的群体的一个重要特征。

需要注意的是，集合{5 6 7 8}并不是一个 LS 集合，因为它不满足 LS 集合的定义。或者说，不满足"该点集内的每个真子集合中存在的到该真子集合在 S

中的补集的关系都多于该真子集合到S外的关系"这个条件。例如,该集合的一个真子集{5 6 7}到补集合{8}的关系有3个,然而真子集{5 6 7}中各个点到集合{5 6 7 8}外的点的关系也有3个,不满足"多于"这个条件。又如,该集合的一个真子集{5 6 8}到补集合{7}的关系有2个,然而真子集{5 6 8}到集合{5 6 7 8}外的关系也有2个,也不满足"多于"这个条件。总之,集合{5 6 7 8}不满足LS集合的定义。

我们还可以这样解释为什么{5 6 7 8}不是一个LS集合。因为对于一个集合来说,只要其中有一个真子集不满足上述条件,该集合就不是LS集合。或者说,一个LS集合中的任何一个真子集与该真子集到其在LS集合的补集上的关系都必须多于到LS集合外的关系。因此,在图6.6中,LS集合{1 2 3 4}包含的真子集合有14个,其中每个到LS集合其余成员之间的关系都多于到非LS集合的成员之间的关系。这一点如表6.7所示。

表6.7 来自图6.6中{1 2 3 4}这个LS集合(该集合的补集是{5 6 7 8 9 10 11 12})中的所有真子集 S_s 及其分别到 N_s-S_s 和 $N-N_s$ 的关系总数

真子集 S_s	N_s-S_s	S_s 到 N_s-S_s 的关系总数	S_s 到 $N-N_s$ 的关系总数
{1}	{2 3 4}	3	0
{2}	{1 3 4}	2	1
{3}	{1 2 4}	3	0
{4}	{1 2 3}	2	1
{1 2}	{3 4}	3	1
{1 3}	{2 4}	4	0
{1 4}	{2 3}	3	1
{2 3}	{1 4}	3	1
{2 4}	{1 3}	4	2
{3 4}	{1 2}	3	1
{1 2 3}	{4}	2	1
{1 2 4}	{3}	3	2
{1 3 4}	{2}	2	1
{2 3 4}	{1}	3	2

利用这种方法,我们可以用表6.8进一步说明集合{5 6 7 8}为什么不是一个LS集合。

表 6.8　来自图 6.6 中{5 6 7 8}这个集合(该集合的补集是{1 2 3 4 9 10 11 12})中的所有真子集 S_s 及其分别到 N_s-S_s 和 $N-N_s$ 的关系总数表

真子集 S_s	N_s-S_s	S_s 到 N_s-S_s 的关系总数	S_s 到 $N-N_s$ 的关系总数
{5}	{6 7 8}	2	1
{6}	{5 7 8}	3	1
{7}	{5 6 8}	2	1
{8}	{5 6 7}	3	0
{5 6}	{7 8}	3	2
{5 7}	{6 8}	4	2
{5 8}	{6 7}	3	1
{6 7}	{5 8}	3	2
{6 8}	{5 7}	4	1
{7 8}	{5 6}	3	1
{5 6 7}	{8}	**3**	**3**
{5 6 8}	{7}	2	2
{5 7 8}	{6}	3	2
{6 7 8}	{5}	**2**	**2**

可见,在集合{5 6 7 8}中,只有{5 6 7}和{6 7 8}这两个真子集不满足"LS 集合"的条件。因为如表 6.8 中的黑体数字所示,{5 6 7}到{8}的关系有 3 个,到除了 5、6、7、8 之外的其他点的关系也有 3 个,前者不大于后者。{6 7 8}这个真子集的情况也一样。因此,集合{5 6 7 8}不是一个 LS 集合。可以想象,如果表 6.8 中第 4 列的黑体数字 3 和 2 分别变为较小的数字,集合{5 6 7 8}就是一个 LS 集合了。

通过上述例示可以看到,对一个图进行 LS 集合的分析是比较麻烦的。特别是对于比较大的网络来说就更复杂了,因为这里要考虑一个集合中的所有可能的真子集。对于一个整体网络来说,如果分析其中一个规模为 N 的点集合是否为一个 LS 集合,则需要考虑该点集中的 $C_N^1+C_N^2+\cdots+C_N^{N-1}=2^N-2$ 个真子集。如果 N 比较小,需要考虑的真子集的数目不多,这还比较容易。可以想象,如果 N 比较大,就要考虑很多真子集了。例如,如果 N 是 5,则需要考虑 $2^5-2=30$ 个真子集。如果 N 是 10,则需要考虑 $2^{10}-2=1\,022$ 个真子集,这是比较复杂的。

在前文中我们已经说明了,一个网络中的任何两个 LS 集合或者没有任何共享的成员,或者一个 LS 集合包含另外一个 LS 集合。换句话说,如果 A 和 B 是一个网络中的 LS 集合的话,那么或者 $A \cap B = \varnothing$,或者 $A \cap B = A$,或者 $A \cap B = B$。A 和 B 之间的关系可能出现的情况只有这三种。因此,我们可以把一个网络中所有可能的 LS 集合构成的集合看成是一系列嵌套的点集合。例如,在图 6.6 中,各个 LS 集合构成如下分区形式(Borgatti, Everett & Shirey, 1990:343):

{{1} {2} {3} {4} {5} {6} {7} {8} {9} {10} {11} {12}}

{{1 2 3 4} {5} {6} {7} {8} {9} {10} {11} {12}}

{{1 2 3 4 5 6 7 8} {9 10 11 12}}

{{1 2 3 4 5 6 7 8 9 10 11 12}}

在 UCINET 软件中并没有给出用来分析 LS 集合的程序,原因之一在于这种程序比较复杂,在笔者看来,更重要的原因在于 LS 集合分析的理论意义不是很大,因为后来学者们(Borgatti, Everett & Shirey, 1990)提出来的 Lambda 集合这个概念已经推广、超越了 LS 集合。因此,Lambda 集合的概念具有明确的理论和实践意义,得到了更多的应用。这就是下面要讨论的内容。

四、Lambda 集合

LS 集合的界定比较严格,它要求其内部成员之间的关系相对比较紧密,但是与外部成员之间的关系相对较少。不幸的是,LS 集合的“这个过剩性质意味着 LS 集合在现实生活中很少出现”(Borgatti, Everett & Shirey, 1990:337)。因此,推广 LS 集合就是必然选择。

(一) Lambda 集合的含义

就关联性而言,一个凝聚子群应该是相对稳健的。也就是说,一个凝聚子群一般不会因为从中拿掉几条线就变成不关联图。正是考虑到这一点,薄加提等学者(Borgatti, Everett & Shirey, 1990)提出了“Lambda 集合”(Lambda set)这个新概念。

这里需要探讨的问题是,在去掉图中的一些线之后,一对点在多大程度上仍然可以通过一条线连在一起?这可以通过这一对点的“边关联度”(line connectivity)指数给出定量测量。一对点 i 和点 j 的“边关联度”指数标记为 λ(i,

j)，它等于为了使得这两个点之间不存在任何路径，必须从图中去掉的线的最小数目。λ(i, j)的值越小，i 和 j 就对去掉的一些线越敏感，即越趋向于分隔开来。反之，λ(i, j)值越大，i 和 j 就对去掉的一些线越不敏感，二者之间的关系越稳健。

现在考察一个拥有点集 N_s 的子图 G_s 中的各个点对(pair of nodes)。如果下面的条件得到满足，我们就称 N_s 是一个 Lambda 集合：N_s 自身内部的任何一对点的边关联度都比由来自 N_s 的一个点和 N_s 外部一点构成的任何一对点的边关联度要大。

从形式化的角度说，一个 Lambda 集合是一个满足如下条件的点集 $N_s \subseteq N$：对于所有的 $i, j, k \in N_s$，并且 $l \in N - N_s$ 来说，都有 $\lambda(i, j) > \lambda(k, l)$ 成立。

Lambda 集合的一个重要性质是，集合中的点就“邻接性”或者“捷径距离”而言不一定要具有凝聚性。一个 Lambda 集合中的成员之间未必邻接，因为在 Lambda 集合中，各个点之间的路径的长度是不加限制的，Lambda 集合的各个点之间可能距离很远。

(二) Lambda 集合分析示例

在 UCINET 中，点击 Network→Subgroups→Lambda Sets 出现如下对话框，找到上述“帮工关系第一子群”矩阵，点击 OK 后，得到该数据中的 Lambda 集合。

表 6.9　Lambda 分析结果

```
LAMBDA SETS
--------------------------------------------------------------------

Input dataset:                    H:\whole UCINET\数据\帮工第一子群

WARNING: data symmetrized by taking larger of Xij, Xji.

HIERARCHICAL LAMBDA SET PARTITIONS

          1 1 1 2 2 2 1 2   1 1 2 1     1           1   1 2
          3 5 4 0 1 2 6 3 9 8 9 4 0 3 5 1 4 6 8 1 2 2 7 7 5

          1 1 1 2 2 2 1 2   1 1 2 1     1           1   1 2
Lambda    3 5 4 0 1 2 6 3 9 8 9 4 0 3 5 1 4 6 8 1 2 2 7 7 5
------    - - - - - - - - - - - - - - - - - - - - - - - - -
    19    . . . . . . . . . . . . . . . . . . . . . . XXX .
    18    . . . . . . . . . . . . . . . . . . . XXXXXXXXX .
    17    . . . . . . . . . . . . . . . . XXXXXXXXXXXXXXX .
    16    . . . . . . . . . . . . . . . XXXXXXXXXXXXXXXXX .
    15    . . . . . . . . . . . . . . XXXXXXXXXXXXXXXXXXX .
    14    . . . . . . . . . . . . . XXXXXXXXXXXXXXXXXXXXXXX
    13    . . . . . . . . . . . . XXXXXXXXXXXXXXXXXXXXXXXXX
    12    . . . . . . . . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    11    . . . . . . . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
    10    . . . . . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     9    . . . . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     8    . . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     5    . . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     4    . XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
     2    XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

分析结果给出了一个图中所有可能存在的 Lambda 集合。第一列给出了多种 Lambda 值,这些数字分别代表不同的“最小边关联度”。对应于不同的最小边关联度就存在不同的 Lambda 集合的分配情况。由图可见,在该数据中存在的 Lambda 集合中 Lambda 最小值是 2,成员包括全体村民,也就是说,如果 Lambda=2,那么全体行动者都可归为一个 Lambda 集合之中。Lambda 的最大值是 19,其成员包含 7 和 17 这两点,如表 6.10 所示。

表 6.10 点对之间的最大流量分析结果

Maximum Flow Between Pairs of Nodes

		13	15	14	20	21	22	16	23	9	18	19	24	10	3	5	11	4	6	8	1	2	12	7	17	25
		13	15	14	20	21	22	16	23	9	18	19	24	10	3	5	11	4	6	8	1	2	12	7	17	25
13	13	25	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
15	15	2	25	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
14	14	2	4	25	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
20	20	2	4	5	25	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
21	21	2	4	5	8	25	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
22	22	2	4	5	8	8	25	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
16	16	2	4	5	8	8	9	25	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
23	23	2	4	5	8	8	9	10	25	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
9	9	2	4	5	8	8	9	10	10	25	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
18	18	2	4	5	8	8	9	10	10	11	25	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
19	19	2	4	5	8	8	9	10	10	11	12	25	12	12	12	12	12	12	12	12	12	12	12	12	12	12
24	24	2	4	5	8	8	9	10	10	11	12	12	25	12	12	12	12	12	12	12	12	12	12	12	12	12
10	10	2	4	5	8	8	9	10	10	11	12	12	12	25	13	13	13	13	13	13	13	13	13	13	13	13
3	3	2	4	5	8	8	9	10	10	11	12	12	12	13	25	14	14	14	14	14	14	14	14	14	14	14
5	5	2	4	5	8	8	9	10	10	11	12	12	12	13	14	25	15	15	15	15	15	15	15	15	15	14
11	11	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	25	16	16	16	16	16	16	16	16	14
4	4	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	25	17	17	17	17	17	17	17	14
6	6	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	25	17	17	17	17	17	17	14
8	8	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	25	17	17	17	17	17	14
1	1	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	17	25	18	18	18	18	14
2	2	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	17	18	25	18	18	18	14
12	12	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	17	18	18	25	18	18	14
7	7	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	17	18	18	18	25	19	14
17	17	2	4	5	8	8	9	10	10	11	12	12	12	13	14	15	16	17	17	17	18	18	18	19	25	14
25	25	2	4	5	8	8	9	10	10	11	12	12	12	13	14	14	14	14	14	14	14	14	14	14	14	25

该程序还计算出任何两个点之间的最大流量(即关联度),并且利用该信息构建 Lambda 集合。上述矩阵就是最大流量矩阵(maximum flow matrix),其中在 i 行 j 列中给出的是 i 和 j 之间的边关联度。对角线的值等于点数。

UCINET 中给出的 Lambda 程序主要分析的是“对称数据”。如果网络数据是有向数据,那么程序将进行对称化处理,其方法是取 x_{ij} 和其对称值 x_{ji} 中的较大者。例如,如果 $x_{ij}=2$,而 $x_{ji}=0$ 的话,那么该程序将令 $x_{ji}=2$。因此,该程序也可用来分析多值对称数据中的 Lambda 集合。

五、社会圈

还有一些学者提出新的凝聚子群概念。在一个图中,不同的 k-派系和 k-丛之间都可能存在一定数目的重叠。进行派系分析(既包括 n-派系,也包括 n-宗派)的结果常常产生一长串重叠的派系,这些结果往往难以解释。一个相对密集的网络常常由大量互相重叠的派系构成。例如,一个包含 20 个点的密集网络可

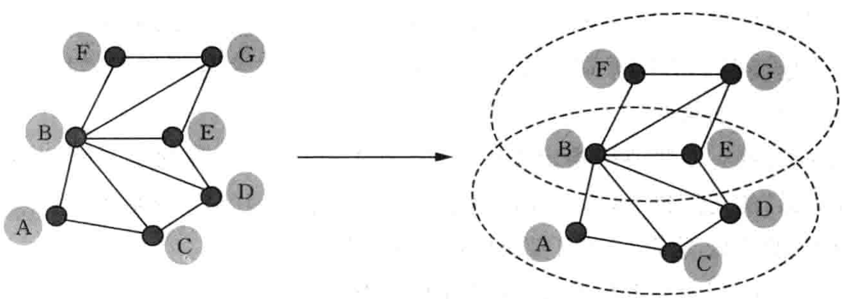
能有 100 个互相重叠的派系。在这种情况下，派系之间重叠的程度可能比派系本身的构成更有意义。因此，社会网络分析者用一些可以明确体现出这种重叠的概念，“社会圈”(social circle)这个概念应运而生(Alba & Kadushin, 1976; Alba & Moore, 1978; Kadushin, 1966)。

“社会圈”这个概念是卡杜山(Kadushin)根据齐美尔(Simmel, 1902)最初的思想提出来的，后者最早提出了“社会圈交叉”的重要性。齐美尔是根据个人网络之间的重叠性和嵌入性来思考社会圈的。一个社会圈的凝聚性不仅仅建立在其成员的“面对面”的接触之上，更建立在较短的间接关系链的基础上，正是这种间接关系链把人们连在一起。

艾尔巴(Alba, 1973)建议，恰当的程序就是应用“滚雪球”方法，这样，派系就会越来越大，成为松散的圈。在分析圈的时候，第一步是找到那些规模为 3 的 1-派系，然后把所有那些只有一个成员不同的派系合并到一个圈中。换句话说，确定圈的第一步标准是，如果各个派系的 2/3 的成员完全相同，则就把这些派系合并为一个圈。这样，经过第一步分析之后，结果是产生了一个或者多个圈、一系列独立的派系以及一些孤立点。

第二步，将其余派系合并到一个圈当中，在这一步中有 1/3 的重叠就可以。这种组合的结果将产生一个大的圈或一系列被联系不很紧密的派系和点所包围的小圈。图 6.7 即对社会圈的一个简单分析。第一步分析出来两个圈，但是它们在第二步被合并在一个圈中(Scott, 2013:117)。

有一点需要指出，即选择什么样的重叠水平标准是没有定论的。研究者必须清醒地知道，他的标准在实际应用中是否有意义。



1-派系:{A、B、C}{B、C、D}{B、D、E}{B、F、G}{B、G、E}
第一步产生的圈:{A、B、C、D、E}{B、F、G、E}
第二步产生的圈:{A、B、C、D、E、F、G}

图 6.7 互相交叉的社会圈

第六节　凝聚子群的分派指数

上文介绍了各种子群的含义及其测量方法,但是没有介绍一个网络在多大程度上可以分成各个子群体。换句话说,上文没有探讨一个网络的分派程度,这就是下面介绍的内容。

一、单类网络中的分派指数

在一个组织中可能存在多个凝聚子群,上一节介绍了各种凝聚子群概念及分析方法。现在的问题是,对于一个整体网络来说,其中存在的分派情况是否严重?除了可以利用上述各种凝聚子群分析方法之外,我们还可以利用一些指数加以测量,E-I 指数就是其一,该指数可用于测量派系林立的程度,甚至可用于危机管理。下面介绍 E-I 指数的含义。

我们可以把网络中存在的关系数量分为两类:派别之间的关系(External links)和派别内部的关系(Internal links)。前者对于组织的危机管理来说往往居于重要地位。据此可以构造一个指数,即 E-I 指数(External-Internal Index, E-I Index)。具体地说, E-I index $=\frac{EL-IL}{EL+IL}$。其中,EL 代表"子群体之间的关系数";IL 代表"子群体内部的关系数"(Krackhardt & Stern, 1988:127)。

实际上,E-I 指数也等于(子群的密度)/(整体的密度)(罗家德,2005:144)。该指数的取值范围为[−1, +1]。该值越向 1 靠近,表明关系越趋向于发生在群体之外,意味着派系林立的程度越小;该值越接近−1,表明子群体之间的关系(即外部关系)越少,关系越趋向于发生在群体之内,意味着派系林立的程度越大;该值越接近 0,表明派别内外关系数量相差不多,关系越趋向于随机分布,看不出派系林立的情形。

该指数拥有一些值得指出来的性质:首先,该指数测量的是"外部关系"对"内部关系"的支配程度,而不仅仅测量了外部关系。因此,该指数不但随着外部关系数量的降低而降低,而且随着内部关系的增加而降低。

克拉克哈特等学者指出(Krackhardt & Stern, 1988:127),之所以用比例的形式计算该指数,主要出于如下三方面考虑。

首先是认同方面。从理论上讲,内部关系的密度越大,子群体内部的成员之间越倾向于相互认同。朋友关系的密度越大,越容易提升整个组织的认同感,因

而可以使得人们成功地面对危机。

其次是成本维度。每个人的时间、精力和需要有限，因此在社会互动和亲密关系的维持方面都是有限度的。超出一定的限度，人们就会考虑是否维持某种子群体内部的关系和子群体之间的关系。在这个意义上可以说，一个人的内部关系越多，用来维持外部关系的精力就越少。因此，内部关系代表了一种子群体内部的“机会成本”。

再次是方法论的考量。之所以在E-I指数中引入内部关系，还有方法论上的考虑。也就是说，在诸如调查“朋友关系”的时候，由于“朋友”的标准不一，因此可能造成调查结果存在误差。这时候，通过比较内部关系和外部关系，就可以较好地控制测量误差的来源，因为诸如此类的测度对调查的误差不敏感。这也是采用此指数的原因。

在UCINET中，沿着Network→Cohesion→E-I index这条路径，就可以分析矩阵的E-I指数。该算法将针对一个矩阵的某种分区计算出E-I指数，并执行一种置换检验，以便评价其显著性。假设一个矩阵被分为各个相互独立的分区，那么根据定义，E-I指数就是群体外的关系数量减去群体内的关系数量，再除以总关系数。虽然该值的取值范围为[−1, 1]，但是对于规模固定的一个网络来说，其取值范围可能是限定性的。该指数也针对每个群体和每个行动者进行计算。同时要进行置换检验，以便检验该网络的E-I指数是否在统计意义上显著地大于或小于某个值。

为了计算E-I指数，需要有两个数据矩阵。一个是关系矩阵，它通常是多值矩阵或二值矩阵；另外一个是属性矩阵，它具体界定了每个行动者隶属于哪个子群体，是有关分区信息的矩阵。也就是说，为了把矩阵分为各个群体，需要指定一个向量，这需要给出数据名、维度（行或列）和一个整数。例如，要利用名为ATTRIB的矩阵的第二行作为分区向量，需要键入“ATTRIB ROW 2”。程序将读取该行的数据，并用这个数据来界定群体。程序将对用来对数据进行分区的属性向量重新编码，计算出与各个群体对应的块密度矩阵。

假设我们分析Camp92数据的E-I指数。需要首先知道该数据中各个点的分区情况，据此构造分区向量并命名为18attribute。然后沿着Network→Cohesion→E-I Index这条路径，点击后在“Input Dataset”一项上选出数据矩阵Camp92，在“Attribute”（分区属性）一项找到刚刚构造的分区向量18attribute的第一列（实际上该矩阵只有一列）。点击OK后即可计算出如下结果。

表 6.11　整体网络的 E-I 指数分析结果

```
E-I INDEX
--------------------------------------------------------------------------------
Adjacency dataset:            H:\Ucinet Data\DataFiles\Camp92
Attribute:                    18attribute
# of Permutations:            5000
Random seed:                  12905
Individual E-I scores:        IndE-I

Warning: This procedure ignores direction of ties.
Warning: Row Attribute vector has been recoded.
Here is a translation table:

Density matrix

               1      2      3      4
               1      2      3      4
          ------ ------ ------ ------
  1 1   10.167 10.250  9.450 10.550
  2 2   10.250 10.333 10.650 10.150
  3 3    9.450 10.650  9.800 10.920
  4 4   10.550 10.150 10.920 12.700

306 ties.

Whole Network Results

                     1       2       3       4
                  Freq     Pct Possibl Density
               ------- ------- ------- -------
  1 Internal    64.000   0.209  64.000   1.000
  2 External   242.000   0.791 242.000   1.000
  3      E-I   178.000   0.582 178.000   0.582

E-I Index: 0.582
Expected value for E-I index is: 0.582

Permutation Test
Number of iterations = 5000
```

计算结果首先给出了四个分区群体的密度矩阵。可见,第一个群体的密度是 10.167,第四个群体的密度最大,达到 12.700。由密度表可以看出,各个群体之间的关系没有表现出明显的模式,例如,群体内部的关系密度表现得不明显。

下面的表给出了整体网分析的结果。第一列给出了“群体内部关系”观察到的频次,第二列计算这些频次与总关系值之比,第三列给出在群体规模已知的情况下可能出现的最大关系值,最后一列是密度指标。

最下端给出了整个网络的 E-I 指数为 0.582。根据前文的讲解,该值越向 1 靠近,意味着派系林立的程度越小。0.582 比较大,说明该整体网络派别林立程度小。

随后又给出了置换检验的结果。第一列表达的是观察值,紧接着的四列依次给出了来自重排检验的最小值、均值、最大值和标准差。随后给出随机检验获得的值大于(或等于)以及小于等于观察值的次数。这些值都用概率的形式表示,可视为显著性水平 p 值。随后的表给出了群体层次的关系和 E-I 指数,这些值都大于 0.5。最后一个表表示每个个体的关系和 E-I 指数,这些值也都比较大,说明没有派系林立情形的出现。

需要指出的是,本例根据的是多值矩阵计算出来的,当然也可以根据二值数据进行分析。

表 6.12 群体层次和个体层次 E-I 指数分析结果

Permutation Test
Number of iterations = 5000

	1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1 Internal	0.209	0.209	0.209	0.209	0.000	1.000	1.000
2 External	0.791	0.791	0.791	0.791	0.000	1.000	1.000
3 E-I	0.582	0.582	0.582	0.582	0.000	1.000	1.000

Group level E-I Index

	1 Intern	2 Extern	3 Total	4 E-I
1 1	12.000	56.000	68.000	0.647
2 2	12.000	56.000	68.000	0.647
3 3	20.000	65.000	85.000	0.529
4 4	20.000	65.000	85.000	0.529

Individual Level E-I Index

	1 Intern	2 Extern	3 Total	4 E-I
1 HOLLY	4.000	13.000	17.000	0.529
2 BRAZEY	3.000	14.000	17.000	0.647
3 CAROL	4.000	13.000	17.000	0.529
4 PAM	4.000	13.000	17.000	0.529
5 PAT	3.000	14.000	17.000	0.647
6 JENNIE	3.000	14.000	17.000	0.647
7 PAULINE	3.000	14.000	17.000	0.647
8 ANN	3.000	14.000	17.000	0.647
9 MICHAEL	4.000	13.000	17.000	0.529
10 BILL	4.000	13.000	17.000	0.529
11 LEE	4.000	13.000	17.000	0.529
12 DON	4.000	13.000	17.000	0.529
13 JOHN	4.000	13.000	17.000	0.529
14 HARRY	3.000	14.000	17.000	0.647
15 GERY	4.000	13.000	17.000	0.529
16 STEVE	4.000	13.000	17.000	0.529
17 BERT	3.000	14.000	17.000	0.647
18 RUSS	3.000	14.000	17.000	0.647

二、多类网络中的分派指数

针对单个网络关系中出现多派的情况，可利用 E-I 指数计算其分派系数。如果在一个网络中存在多种关系，或者各个行动者都有自己的属性，如何探讨其分派情况，进而计算“分派”(segregation)指数呢？弗里曼(Freeman，1972)认为，从理论上讲，对于多种社会关系来说，如果已知群体成员的某种属性(如性别、性格、贫富等)在网络中没有明显的影响力的话，那么至少就这种属性来说，可以把关系看成是随机分布的，因而看不出来一个网络中是否出现严重的分派现象，看不出网络分为多少个群体。因此，在理论上期望出现的群体之间的关系数(number of cross-group ties)和实际出现的群体之间的关系数就测量了“分派指数”(segregation)。其计算公式具体可表达为：

$$Seg = \frac{E(X) - X}{E(X)}$$

其中,$E(X)$代表群体之间期望出现的关系数,X 代表群体之间实际出现的关系数。下面用一个假设的例子展示具体的计算过程。

在一所高中的一个班级中,同学之间存在着人际关系网络(日常往来),每一位同学又有两类属性:性别和性格。性别有两种:男和女(在图 6.8 中分别用深色和浅色表示,例如,1, 2, 3, 4, 5, 6 的颜色是深色,表明他们都是"男生";其余点的颜色是浅色,表明她们都是"女生");性格也有两种:内向(用圆形表示,由图中可见,圆形有 4, 5, 6, 7, 8, 9,表明这些点都是"内向型性格"的学生)和外向(用方形表示,由图中可见,方形有 1, 2, 3, 10, 11, 12, 13, 14, 15,表明这些点都是"外向型性格"的学生)。

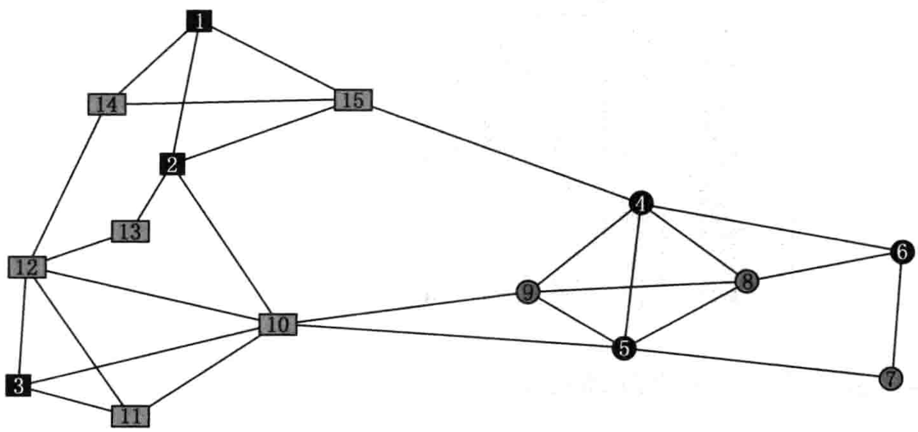


图 6.8　15 个同学之间的人际关系网(性别和性格)

要测量"分派指数",首先要计算数据中群体之间实际出现的关系数,然后计算群体之间期望出现的关系数。这需要先创建混合矩阵(mixing matrix),据此计算不同性别的人之间的交流次数,不同性格的人之间的交往次数等,最后计算出分派指数①。当然,在进行这种计算工作之前,首先要把网络数据输入到 UCINET 之中。可以按照如下数据语言格式输入(当然还有其他格式):

① 参见美国杜克大学穆迪(Moody)教授的个人主页:http://www.soc.duke.edu/~jmoody77/s884/index.htm。

```
DL n=15
format = nodelist1
data:
1   2 14 15
2   1 10 13 15
3   10 11 12
4   15 9 5 8 6
5   10 9 4 8 7
6   4 7 8
7   5 6
8   4 5 6 9
9   10 4 8 5
10   11 3 12 2 5 9
11   3 12 10
12   14 13 10 11 3
13   2 12
14   1 15 12
15   1 14 2 4
```

然后把该数据语言转换为UCINET数据，转换方法是在UCINET中沿着Data→Import Text File→DL路径，选择该文本文件A，即可转换为数据矩阵，可以把该数据矩阵文件也命名为A。

（一）构造标示矩阵

下一步，在UCINET中分别构造两个标示矩阵(indicator matrix)：性别标示矩阵和性格标示矩阵。从图6.8中可以看到，第1到第6个人都是"男性"，因此，在性别标示矩阵中，"男"这一行下的前六列数字都是"1"；第7到第15个人都是"女性"，因此，在标示矩阵中，"女"这一行的前7到15列都是"1"。如表6.13所示。

表6.13 性别标示矩阵

男	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
女	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

同理可构造"性格"标示矩阵，如表6.14所示。

表6.14 性格标示矩阵

内向	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0
外向	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1

上述两个标示矩阵的构造需要用到UCINET中的"数据表"(spreadsheet)，即点击Data→Data Editors→Matrix Editor之后，即可打开数据表格式，按照上述表示输入这两个矩阵，可分别命名为"sex"和"character"。

（二）转置

为了计算混合矩阵M，需要把标示矩阵乘以邻接矩阵，再乘以标示矩阵的转置矩阵。用矩阵语言表示就是，混合矩阵M=ΓAI。

$$\underset{(k\times k)}{M} = \underset{(k\times n)}{\Gamma}\ \underset{(n\times n)}{A}\ \underset{(n\times k)}{I}$$

因此，需要首先利用 UCINET 中的“转置”(transpose)命令(或利用 UCINET 中的 Matrix Algbra)，把标示矩阵进行转置处理，得到该矩阵的转置矩阵(在 UCINET 中，沿着 Data→Transpose 这条路径，分别选择上述 sex 文件和 character 文件，点击 OK 后即可得到其转置矩阵，输出结果分别自动命名为 sex-Transp 和 character-Transp)。

混合矩阵的计算要用到 UCINET 中的矩阵算法。具体来说，在 UCINET 中，沿着 Tools→Command Line/Matrix Algebra 这条路径，点击之后出现对话框。在 type command below 对话框中输入如下内容：

b=prod(sex, A)

其含义为计算 sex 和 A 之积，将乘积矩阵命名为 b，回车。再键入 M=prod(b, sex-Transp)(即计算 b 和 sex-Transp 这两个矩阵之积，将乘积矩阵命名为 M)，回车。在 type command below 对话框中再输入“Disp M”，即展示(display)上述计算得到的矩阵 M，回车之后就得到计算出来的混合矩阵 1。同理可计算得到混合矩阵 2。经过这种矩阵计算和分析，得到如下两类混合矩阵。

表 6.15　两种混合矩阵

混合矩阵 1		
	男	女
男	6	17
女	17	16

混合矩阵 2		
	外向	内向
外向	20	3
内向	3	30

(三) 分派指数(segregation index)的计算

为了计算期望值，需要利用列联表中期望频次的标准计算公式，即期望值=行边缘值×列边缘值/总和值，得到的结果如表 6.16 所示。

表 6.16　性别观察值与期望值矩阵观察到的值及期望值(括号内为期望值)

	男	女	行边缘和
男	6(9.45)	17(13.55)	23
女	17(13.55)	16(19.45)	33
列边缘和	23	33	56

由于期望值 E(X)是"群体之间的关系总和",因此它等于(13.55＋13.55);实际值 X＝(17＋17)。所以,分派指数 Seg＝(27.1－34)/27.1＝－6.9/27.1＝－0.25。分派指数的计算可以手工进行,当然也可以编写小程序进行计算。

对于"性格"这个属性来说,观察到的值及期望值(括号内为期望值),如表 6.17 所示。

表 6.17　性格观察值与期望值矩阵

	内　向	外　向	行边缘和
内　向	20(9.45)	3(13.55)	23
外　向	3(13.55)	30(19.45)	33
列边缘和	23	33	56

可见,E(X)＝(13.55＋13.55);X＝(3＋3)＝6;所以,分派指数 Seg＝(27.1－6)/27.1＝21.1/27.1＝0.78。

(四) 分派指数的现实意义及存在的问题

下面分析分派指数的现实含义。上文已经指出,分派指数的计算公式为 $Seg=\frac{E(X)-X}{E(X)}$。E(X)代表期望出现的群体之间的关系数量,X 代表实际出现的群体之间的关系数量。现在的问题是,分派指数的区分范围是什么? 它具有怎样的含义? 以表 6.18 为例。

表 6.18　性别之间的关系矩阵

	男	女
男	a	b
女	c	d

(1) 如果 a＝d＝0,计算出来的分派指数将达到最小值－1,意味着在每个子群体的内部不存在关系,关系完全存在于各个子群体之间。这时候不但不存在分派,而且关系完全出现在各个子群之间;

(2) 如果 a/c＝b/d 的话,计算出来的分派指数将等于 0,意味着不存在分派;

(3) 如果 b＝c＝0,那么计算得到的分派指数将达到最大值 1,意味着关系完全存在于每个子群体内部,在每个子群体之间不存在关系,整个群体分隔为相互独立的多个子群体,此时整个网络的分派程度达到最大。从公式中可以看出,

此时 X 等于 0,也就是不出现群体之间的关系,分派指数达到最大值 1。上述数据中的"性格"分派指数 Seg＝0.78,这个值比较大,说明同学们之间的人际关系网在"性格"方面存在着较大的隔离现象。"性别"方面的分派指数 Seg＝－0.25,说明不同性别的同学之间的人际关系没有表现出较大的隔离现象。

计算该分派指数的目的是为了用一个指标说明一个整体网络中存在的分派情况。在某些情况下该指数是很有用处的。

(五)对分派指数的推广

分派指数有一个不好的性质,毋宁说这是该指数的一个问题,即它不是"边缘无关"的(margin free)。也就是说,如果改变感兴趣的属性(如性格)的分布(比如乘以一个常数)而不改变"友谊"和"性别"之间的本质关系的话,就会得到一种不同的分派指数,这也是在列联表中存在的问题。在列联表中,相对频次相同的列联表,在统计检验中其显著性并不相同(卢淑华,2005:285)。当样本容量增加 k 倍,如果相对频次不变的话,那么卡方值将增加 k 倍,这也意味着分派指数将增加 k 倍。

解决这个问题的一个办法就是利用发生比率或比值比(odds ratios)。在该例子中,比值比分析会告诉我们,处于同一种属性的人之间互选对方为朋友的相对可能性(relative likelihood)。

比值比的含义是,处于同一群体的人互选对方的可能性是多少。我们可以根据一个群体中的关系数量及其相对规模来计算比值比,可用表 6.19 所示的表。

表 6.19　群体成员之间的朋友矩阵

	同一群体成员	不同群体成员
朋　友	A	B
非朋友	C	D

比值比 OR＝AD/BC。对于上述数据网络来说,其中有 6 个外向之人,9 个内向之人。他们之间实际出现的关系以及在理论上可能出现的最多关系数量分别如表 6.20 所示。

表 6.20　不同性格的人之间的关系矩阵

	内向	外向
内向	20	3
外向	3	30

	内向	外向
内向	30	54
外向	54	72

可见，在“内向”群体中出现的关系有20个，在“外向”群体中出现的关系有30个，总结在一起，发生在“同一类群体”中的关系数量为20＋30＝50。在“内向”群体中没有出现的关系有30－20＝10个，在“外向”群体中没有出现的关系有72－30＝42个，总结在一起，发生在“同一类群体”中的关系数量为10＋42＝52。在“内向”群体和“外向”群体之间出现的关系有3＋3＝6个，没有出现的关系有54＋54－6＝102个。这种分析可总结如表6.21所示。

表6.21　朋友隶属于群体的关系矩阵

		群　体	
		同一群体	不同群体
朋友	是	50	6
	不是	52	102

因此，比值比OR＝(50)102/52(6)＝16.35。这是一个比较大的值，说明在该群体中存在较大的分派性。当然，比值比不是一个标准化的量纲，有时候难以看出分派程度的大小。

第七节　凝聚子群分析示例

派系、n-派系、成分、n-宗派、k-丛、k-核、LS集合、Lambda集合、社会圈等都属于“凝聚子群”范畴，都可以看成是“凝聚子群分析”(cohesive subgroup analysis)的各个类型。如果网络的规模较小，分析这些子群的步骤就不是很复杂。但是，当网络规模较大，分析这些概念的方法、技术和步骤都很复杂时，就离不开电脑程序了。随着网络规模的增加，计算所需要的时间呈指数增长。分析凝聚子群的一些算法都要搜索数据，发现问题的解。搜索的时间可能随着网络规模的增加而急速增长。一般来讲，大网络的关系趋于稀疏而不紧密，因此，边的数量不是一个重要的问题。对于包含几百个或几千个点的网络来说，只要其中没有太多的群体，派系分析就是可以进行的。对于派系和k-宗派研究来说，边的数目和派系的数目是需要考虑的。

总的来说，大网络的密度一般不大，在计算的时候，可能出现成百上千个派系。在找出派系的时候，如果出现计算上的困难，分析者可以考虑增加派系的规

模,这将大大减少分析出来的凝聚子群的数目。也就是说,如果在计算方面有问题,那么分析者应该考虑逐渐增加一个群体的最小规模。在搜索群体的时候,计算量是很大的。如果说对于拥有 100 个点的网络来说可以做到子群分析的话,那么分析 1 000 个行动者的 15 个群体的计算量甚至计算机都无能为力。因此,对于非常大的网络来说,可以考虑分析之前把该网络分解为比较小的部分,把网络分成各个子群、位置(positions)、"块"或者成分。

因此,在分析凝聚子群的时候,需要经过一些依次递进的步骤,这些步骤基本上都可以在 UCINET 中得到分析(本部分内容参考了 Everett, 2002:102—109)。总体而言,这些步骤坚持的分析逻辑是:始于界定松散、易于寻找的子图,逐渐分析出界定严格、难以发现的子图。当然,在具体分析的时候,研究者一定要结合自己的研究兴趣、针对具体数据的实质、问题的性质等进行凝聚子群分析。分析过程未必严格遵循这里介绍的顺序。

一、进行凝聚子群分析的几个指导性步骤

第一步,如果数据是二值的,直接看第二步。如果数据是多值的,可以有两种处理方法。

一是利用多维量表(Multi-Dimensional Scaling, MDS)(在 UCINET 中点击 Tools→Scaling/Decomposition→Metric MDS 这条路径)或者层次聚类方法(沿着 Tools→Cluster Analysis 这条路径)(假设数据是对称的)进行分析。

二是对数据进行"二值化"(dichotomize)处理(Transform→Dichotomize),但是要注意数据的性质。在某些研究中,在收集和编码初始数据的时候,数字越大表示关系越近,这样的数据被称为"相似性"(similarity)数据。要对相似性数据进行二值化处理,应确保大于某个临界值的数字重新编码为"1",否则为"0"。如果数据是"相异性"(dissimilarity)数据(即距离数据,数字越大表示关系越远),在二值化处理的时候则相反,即把大于某个临界值的数字重新编码为"0",否则为"1"。可以根据不同的临界值进行分析,以便检验数据的稳健性,找出数据中隐含的比较完备的结构。

如果关系是多元的(例如,在一个行动者群体中同时存在多类关系,如"朋友关系"、"邻居关系"和"亲属关系"等),则需要分别进行处理和分析。可以首先对数据进行开包(在 UCINET 中沿着 Data→Unpack 这条路径),然后分别对各个

数据进行分析。有的程序可以直接针对多维数据矩阵进行分析（例如分析“块”的程序，这时候就不用“开包”分析）。

第二步，分析“成分”。对于有向关系数据来说，既需要找出“强成分”（strong components），也要找出“弱成分”（weak components）（在 UCINET 中，沿着 Network→Regions→Components 这条路径）。我们不但可以分析二值关系网络中的成分，也可以分析多值关系网络中的成分。

成分分析是凝聚子群分析的最简单形式，有时候成分分析可以为我们提供用来回答问题的充分信息。如果情况确实如此，分析就到此结束。如果成分分析的结果并没有为我们提供充分的信息，那么，对于无向关系数据来说，需要进行第三步分析；对于有向数据来说，要首先进行对称化处理（symmetrize）。

第三步，如果需要分区，并且已经知道可以分为几个区，可以直接进行下面第六步的分析。否则先进行派系分析，找出所有的派系（Network→Subgroups→Cliques）。这里又分为几种情况。

1. 如果在分析结果中没有找到派系或者仅仅找到几个派系，则可以试着进行如下分析：

（1）如果派系的规模最小是 4，可以降低规模（但是不能降低到 3 以下）；

（2）如果数据是对称的，则需要结合原始数据分析是否真的没有派系或者有很少的派系。如果原始数据中的行动者之间的关系很少，那么存在的派系很可能在现实中确实较少，分析到此结束；

（3）如果数据经过了二值化处理，则对于相似性数据来说，需要降低临界值，对于相异性数据来说，需要提高临界值。如果所有这些努力都不可行，则按照如下第五步进行分析，即找出其所有的 k-丛。

2. 如果分析的结果是找到了太多的派系（这只有在第四步之后才可明显出现），则把上述三个步骤颠倒过来进行如下分析。首先，如果没有什么互惠关系可以改动的话，则需要增加派系的最低规模（这一步总是比较有效的），最终也就是改变临界值。如果分析的最终结果是出现了一个比较简单的派系序列，则可以结束分析了。但是，除非从最终的结果中可以看出很简单的结构，否则需要进行第四步分析。

第四步，分析派系重叠的模式。如果存在大量派系，派系之间就会有很多重

叠。这种重叠性可能隐藏了派系的结构。利用“共享成员”(co-membership)矩阵方法(参见本书第三章)可降低派系的数量,目的是找出派系中的主要成员、局外人(outsiders)以及群体中可能存在的领导者。通过这个步骤,我们也可能推演出网络中大体上有多少个分组,这个信息将有助于我们进行“分派”(factions)分析。或者说,如果已经大体上知道网络中可以分为多少个小派别,则可以进行分派分析,即第六步的分析,否则分析结束。

第五步,找出所有的 k-丛(k-plexes),这一步要在上述派系分析的步骤失败之后再进行,因为这时候可以看出,在要分析的数据中行动者之间的联系比较松散。如果情况并非如此,并且几乎不存在什么 k-丛,那么需要进行第三步分析。如果这一步也失败,就需要增加 k 的值,重新进行分析。只是要注意的问题是,k 的值不能太大。关于 k 值的大小和网络规模之间的限制性关系,前文已经给出一个经验公式,可以作为参照。图 6.9 是在点击 UCINET 中的 K-Plex 之后出现的对话框。研究者只要选出需要加以分析的输入数据,即可找到该网络中存在多少个 k-丛。当然,这首先要研究者自己确定 k 的取值是多大以及 k-丛的最小规模。

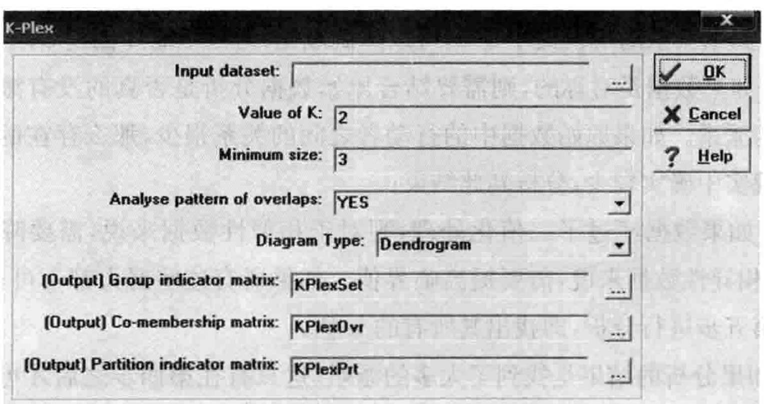


图 6.9 k-丛对话框

如果这种分析仍然无效,则需要转向第六步分析。如果分析的结果是找到了一个简单的 k-丛列表,分析到此结束。但是,与分析派系一样,艾弗雷特(Everett, 2002)建议进一步分析 k-丛之间的重叠性,即将 k-丛看成是派系,从而像在第四步那样进行分析。

第六步，分派分析，找出指定数目的派别（factions）。针对网络中的任何一个成分，都可以进行分派分析，因为，各个独立的成分总是归属于各个不同的派别之中。因此，只有那些需要进一步分成各个较小子群的成分才需要进行分派分析。我们首先需要指定派系的数目，然后找出各个派系的成员（Network→Subgroups→n-cliques 或者 n-clan 或者 k-plex 等）。在 UCINET 中，"faction"命令指的是分析出指定数目的派系情况，而"cliques"命令则无此功能。由于不同性质的网络具有不同的子群性，并且子群内部还可能分出子—子群，因此，我们需要结合具体的关系网络进行检验。

另外，在 UCINET 中，也可以通过分析找到一个网络图中的"块"大约有多少（沿着 transform→Aggregate（include CSS）→block 路径进行分析即可）。例如，对于上文分析的 WOLF-WOLFN 数据来说，可以分析四个年龄段小组的"块"密度。为此，要构造一个列向量表示四个年龄组。在前文构造的 WOLF-WOLFN_GT_6 age 数据中，只需要将第一列的值改变一下即可。例如，第 2，3，8，12，16 只猴子是第 3 组的成员，则在第 2，3，8，12，16 这 5 行第 1 列的值都修改为 3 即可。修改后的数据文件可命名为 WOLF-WOLFN_GT_6 age1。然后依次点击 transform→Aggregate（include CSS）→block 路径，在"Input dataset"项目右侧选出 WOLF-WOLFN 数据，在 Column partition/blocking（if any）右侧选出 WOLF-WOLFN_GT_6 age1 数据，对话框会自动对第一列（Col 1）进行分区分析。如图 6.10 所示。

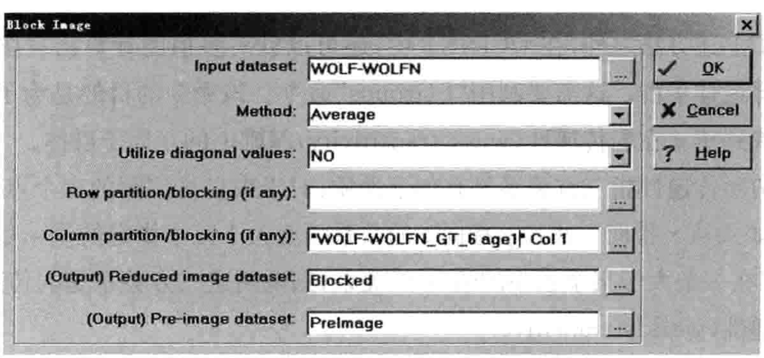


图 6.10　Block Image 对话框

点击 OK 后，得到如下部分结果。

表 6.22 块分析结果

```
NETWORK BLOCK DENSITIES
------------------------------------------------------------
Method:                 Average
Input dataset:          WOLF-WOLFN(D:\datafiles\WOLF-WOLFN)
Column
Block  Members:
 -----  ---------------
    1:  7 14 18 20
    2:  4 5 9 10 15 17
    3:  2 3 8 12 16
    4:  1 6 11 13 19

Reduced Block Matrix
             1      2      3      4
          -----  -----  -----  -----
    1     3.329  3.956  4.442  3.705
```

可见,"块分析"会计算出每一个子群体的密度,计算方法要用到两项之商。第一项是每一个子群体的所有成员实际存在的全部关系数,第二项是子群体的规模乘以(整体网络规模−1)。前者除以后者得到的商就是每个子群体的密度。例如,整个网络规模为 20,第一个子群体的规模是 4。所以,第一个子群体的密度等于这四个成员实际存在的关系数 253 除以(20 − 1) * 4,即 253/((20 − 1) * 4) = 3.329。其余子群体密度的计算与之类似。关于利用上述思路分析具体数据的例子请具体参见(Everett, 2002:105—109),此处不再赘述。

除了上述方法之外,在 UCINET 中,还可以针对多值图计算出在图论意义上的全部最佳子群。这需要利用"f-Groups"命令。该命令的目的是为了在多值矩阵中找到建立在弱传递性(weak transitivity)基础上的互斥子群体。

所谓弱传递性指三方关系具有如下性质:对于来自多值图的三个点 x、y 和 z 来说,如果从 x 指向 y 并且 y 指向 z 的关系强度大于某个指定的值 s,就在 x 和 z 之间存在一条大于某个较小的值 w 的关系,则说明这三方关系是弱传递的,具有弱传递性(weakly transitive)。

"f-Groups"命令将取图中最大值为 s,令使用者事先指定的值为 w,进而检查弱传递性(weak transitivity)。s 的值逐渐降低,直到发现一个违背"弱传递性"条件的三方组存在。然后可利用 s 值对图进行二值化处理。在这种二值化

的图中找到的各个成分(对于有向图来说是弱成分)将形成各个互斥的 F-群体(F-groups)。需要注意的是,这种命令往往分析的是多值图。

二、凝聚子群分析示例

下面我们结合博加提(参见 Steve Borgatti 的个人主页)的例子,用一个完整的实例展示凝聚子群分析的各个步骤。1956 年,纽科姆(Newcomb)对密西根大学相互不熟悉的 17 名交换生(本科生)之间的友谊关系进行了实验研究。在 16 周膳宿时间内,他们相互提名,即根据“感觉喜欢”对其他 16 名学生喜欢的等级进行排序。下面分析的是第 15 周的数据,等级值最小为 1,表示最喜欢;最大值为 16,表示最不喜欢。在此数据中也允许存在“无关”的情况(即不作评价)。该数据如表 6.23 所示。

表 6.23 17 名交换生之间的友谊关系

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
1	0	12	15	5	10	11	6	4	7	16	8	9	2	3	13	14	1
2	8	0	13	2	3	6	9	10	5	15	7	4	11	12	14	16	1
3	8	11	0	10	12	3	5	13	4	14	6	2	9	15	7	16	1
4	6	4	15	0	3	2	10	11	5	16	9	8	7	14	12	13	1
5	5	4	13	2	0	8	10	6	1	14	12	11	3	9	15	16	7
6	6	9	14	3	8	0	7	1	2	15	13	11	4	10	12	16	5
7	12	4	8	6	14	10	0	5	9	16	2	1	7	11	13	15	3
8	1	9	15	3	6	4	13	0	11	14	10	8	2	7	12	16	5
9	10	5	13	3	7	1	12	9	0	16	11	6	8	4	14	15	2
10	2	12	14	11	10	6	3	4	7	0	9	1	15	13	5	16	8
11	9	3	6	4	7	13	5	14	8	16	0	2	10	11	12	15	1
12	8	2	12	7	11	14	1	10	3	16	5	0	6	9	15	13	4
13	1	10	14	9	8	5	3	2	7	15	12	11	0	6	13	16	4
14	4	9	16	10	15	2	8	11	1	14	3	7	6	0	12	13	5
15	12	8	11	3	16	7	9	13	4	14	15	5	6	10	0	2	1
16	12	5	16	3	11	8	7	15	2	14	9	1	13	10	6	0	4
17	4	3	14	2	6	10	9	11	1	16	8	7	5	13	12	15	0

第一步:二值化处理。显然,该数据不是二值数据,因而在具体分析之前需要进行二值化转换(当然可以对多值数据进行分析,这里暂且不考虑这些)。其原因在于,首先,就两个行动者之间的比较来说,等级秩序是有问题的。具体地说,如果某人喜欢所有人,另外一个人都不喜欢其他人的话,把每个人都放置在

一个等级之中就不恰当。其次,有时候用不着把等级值分为16个等级,因为这种分类太细了。因此,在某些情况下我们建议不使用等级值。

我们可以根据自己的想象或者事先得到的研究经验,假定一个人的前五项选择是重要的,因此,把全部等级值为1到5的选项都用1来代替,所有其他值都用0来代替。具体的转换步骤为:假设上述矩阵名为"17like",并且该矩阵位于桌面上,那么在UCINET中,点击Transform→Dichotomize后会出现对话框,如图6.11所示。点击"GT-Greater Than",会下拉出现五个选项,选择LT,即"小于"项后,再把"Cut-Off Value"选项上键入"6"(因为我们要求小于6的值用"1"来代替)。在此矩阵中,对角线是不予考虑的。因此,可使"Diagonal OK"一项保持不变,仍然为"No"(如果对角线予以考虑,则需要将"Diagonal OK"一项的"No"变为"Yes")。

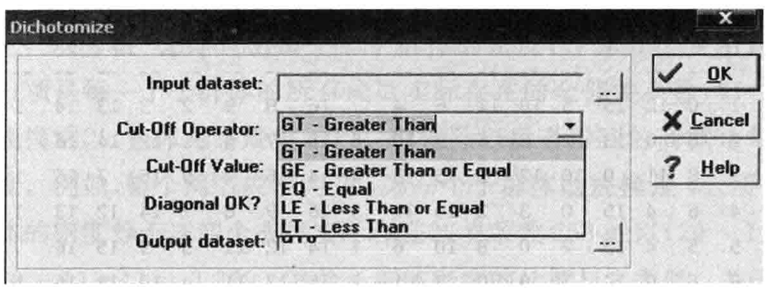


图6.11 "二值化"对话框

在"Output dataset"(即输出数据)选项上,需要研究者自己确定输出的结果放在哪个磁盘里,并且可以起新的名字,本例子的输出文件名为"LT6",如图6.12所示。

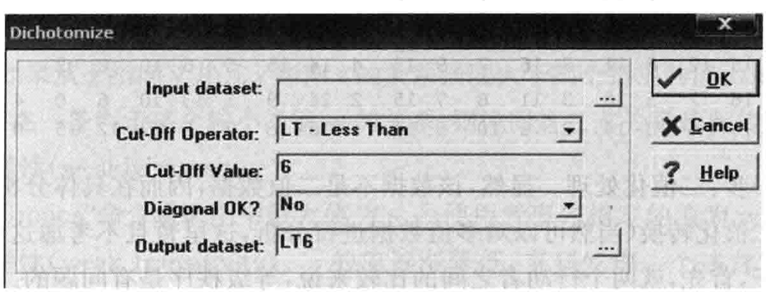


图6.12 "二值化"对话框

经过这种转换得到的矩阵如下：

表 6.24　二值化转换后的数据

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	1
2	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1
3	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	1
4	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1
5	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
6	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	1
7	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1
8	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1
9	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1
10	1	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0
11	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1
12	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1
13	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1
14	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	1
15	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1
16	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1
17	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0

第二步:进行成分分析。成分分析有两类:弱成分分析和强成分分析。成分分析过程如下。在 UCINET 中,点击 Network→Regions→Components→Simple graphs 后出现对话框,如图 6.13 所示。其中“Kind of components”一项有两种选择,即“弱”(weak)和“强”(strong)之分。

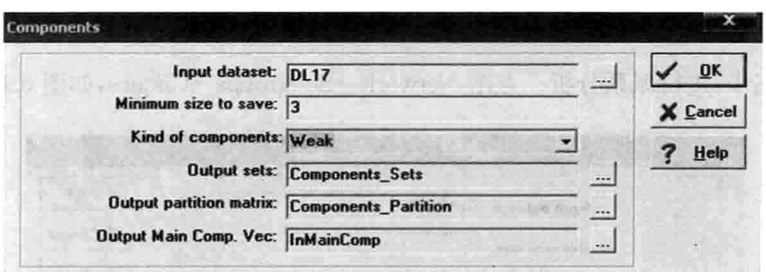


图 6.13　“成分”分析对话框

点击“Weak”一项,选择 OK 即可得到分析结果,发现该数据只有一个弱成分,是由全部行动者构成的。如果点击“Weak”右侧的箭头,即可出现“strong”选项。选择 OK 即可得到进行强成分分析的结果,发现该数据包含五个强成分,即存在一个大的强成分(由 1, 2, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 17 这 13 个

点构成)和四个较小的强成分(分别由 3，10，15 和 16 这四个单点组成)。

这种分析还是没有为我们提供有关该群体内部结构的任何洞见。因此,我们需要对该数据矩阵进行对称化处理,继续进行更细致的分析。这需要点击 Transform→Symmetrize,其中“Symmetrizing Method”一项可以有 14 种选法,在此我们利用互惠关系对该矩阵进行对称化处理,也就是说,只有满足 $x_{ij}=x_{ji}=1$ 的矩阵元素才令它为 1,否则都为 0。因此,我们选择“Symmetrizing Method”中的“Minimum”,因为这项指令把矩阵中的各个值 x_{ij} 和 x_{ji} 的值都用二者之中较小者来代替($i<j$),这就符合我们所说的“互惠性”这个条件。按照这种方法进行处理得到如下对称矩阵:

表 6.25 交换生数据的对称化处理

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	1.000
2	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	1.000	0.000	0.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
5	0.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000
8	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
9	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
12	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
13	1.000	0.000	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
14	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
17	1.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000

Symmetrized matrix saved as dataset C:\Program Files\Ucinet 6\DataFiles\SymDL17

第三步:进行派系分析。点击 Network→Subgroups→Cliques,如图 6.14 所示。

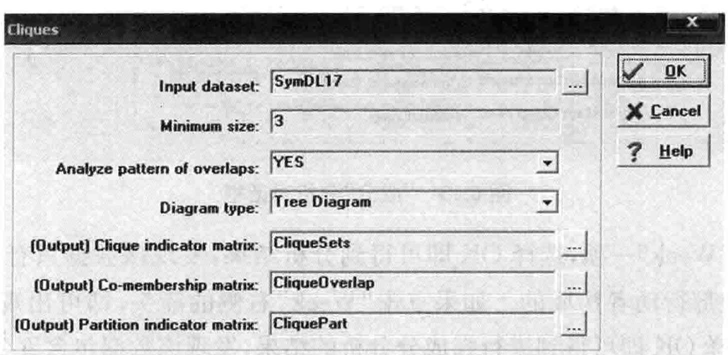


图 6.14 派系分析对话框

得到如下七个派系：

表 6.26　派系分析结果

```
1:  2   4   9  17
2:  2   4   5
3:  1   8  13
4:  1  13  17
5:  4   6   9
6:  6   8  13
7:  7  11  12
```

Actor-by-Actor Clique Co-Membership Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	2	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	1
2	0	2	0	2	1	0	0	0	1	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	2	0	3	1	1	0	0	2	0	0	0	0	0	0	0	1
5	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	1	0	2	0	1	1	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
8	1	0	0	0	0	1	0	2	0	0	0	0	2	0	0	0	0
9	0	1	0	2	0	1	0	0	2	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
12	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
13	2	0	0	0	0	1	0	2	0	0	0	0	3	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	2

HIERARCHICAL CLUSTERING OF EQUIVALENCE MATRIX

```
            1   1 1 1 1 1         1       1
Level     3 0 7 1 2 4 5 6 5 6 8 1 3 2 4 9 7
-----     - - - - - - - - - - - - - - - - -
2.000     . . . . . . . . . . . XXX XXX . .
1.667     . . . . . . . . . . XXXXX XXXXX .
1.000     . . XXXXX . . . . . XXXXX XXXXXXX
0.750     . . XXXXX . . . . XXXXXXX XXXXXXX
```

```
0.417   . . XXXXX . . . .   XXXXXXXXXXXXXXX
0.114   . . XXXXX . . .   XXXXXXXXXXXXXXXXX
0.000     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

Clique-by-Clique Co-membership matrix

	1	2	3	4	5	6	7
	-	-	-	-	-	-	-
1	8	2	0	1	2	0	0
2	2	6	0	0	1	0	0
3	0	0	6	2	0	2	0
4	1	0	2	6	0	1	0
5	2	1	0	0	6	1	0
6	0	0	2	1	1	6	0
7	0	0	0	0	0	0	6

从中可以看到,3, 10, 14, 15 和 16 这五人不隶属于任何派系,他们是孤立者。显然,这种派系分析告诉我们的信息是比较多的。我们可以更深入地分析。

第四步:群体共享成员分析。群体共享成员(group co-membership method)方法产生一个包含如下成员的大群体{1, 2, 4, 5, 6, 8, 9, 13, 17}和小群体{7, 11, 12} 以及一些局外人士。共享群体聚类分析与这种分析的结果是一致的,但是新提供了关于大网络结构的洞见。大网络可分为{2, 4, 5, 6, 9, 17}和{1, 6, 8, 13, 17}这两种子网络,这表明行动者 6 和行动者 17 在网络中居于重要地位。

在本例中,所汇报的群体来源于成分分析,不必进行派系分析。如果指定进行三派的分派分析,计算出来的分组结果是{1, 2, 4, 5, 6, 8, 9, 13, 14, 17}, {10, 15, 16}和{3, 7, 11, 12},这与前述的分析结果很一致。在本例中,{10, 15, 16}群体与其他群体之间无关联。由此得出的结论是,上述分析是稳健的,能够代表数据的内在结构,分析结束。

本 章 小 结

凝聚子群是一种含义广泛的子群概念,其研究的目的是揭示群体内部的子结构。本章介绍了多种凝聚子群概念。派系:如果该子图是规模不小于 3 的最大完备子图。n-派系:该子图中的任何两点之间(在总图中)的距离都不超过 n。

n-宗派:子图中任意两点之间在子群内部的捷径距离不超过 n。k-丛:子图中任何点与至少 n-k 个其他点相连接。k-核:子图中任何点与至少 k 个其他点相连接。成分:子图中任何两点都可以通过一定的途径相连。Lamda 集合:子图 Ns 内部的任何一对点的边关联度都比任何一个由来自 Ns 的一个点和 Ns 外部一点构成的点对的边关联度要大,这样的子图 Ns 就叫做 Lamda 集合。LS 集合:假设总图 G 对应的点集为 N;并且 G 中存在一个子图 Gs,其对应的点集记做 Ns;再假设 Ns 中存在一个真子集合 Ss,即 Ss⊂Ns。如果任何一个真集合 Ss 到 Ns-Ss(即 Ss 在 Ns 中的补集)中点的关系都多于到 N-Ns 中点的关系的话,则称 Ns 是一个 LS 集合。在无向网络、有向网络和多值网络中,上述每种凝聚子群的含义又有所改变。

介绍了单类网络、多类网络中分派指数的计算,目的是用一种指数测量群体的分派程度。

在分析凝聚子群的时候,可以坚持如下步骤。首先分析定义比较严格的凝聚子群,然后分析界定比较松散的子群。例如,可以先分析“派系”,如果不存在派系,应该进一步分析 n-派系、n-宗派、k-丛、k-核、成分、Lambda 集合等。如果在研究之初就知道一个整体网分为几个派系,但是不能确定其成员的话,则可直接利用 faction 程序进行分析。

找到凝聚子群之后,对它们的解释也离不开这些凝聚子群本身具有的属性特征。因此,要结合“属性资料”对凝聚子群进行解释和说明。

第七章

社会网络的关联性与凝聚度

第六章研究了凝聚子群的各种类型，并举例展示了凝聚子群的分析步骤。这种分析只是将整体网络中的某些联络紧密的子群体区分出来，并没有严格地界定什么是凝聚力。除了凝聚力之外，在实际问题中，我们往往还关注网络的另外一类特点，这就是关联性。关联性与凝聚性是不同的。一个网络可能关联度大，但是凝聚力未必大。如果再结合现实世界中人们对资源的争夺，可以更加清楚地看到这一点，例如，一个大的网络可能包含几个小网络，每个小网络内部联系可能非常紧密，但是整个网络可能要分崩离析。对于现实世界中的近亲之间的关系来讲，即便他们之间的联系未必紧密，但是他们往往仍然有凝聚力。

凝聚力问题与社会学的一些经典问题是联系在一起的。例如，凝聚力与经典社会学中的"社会团结"研究有关。从社会网络的角度上讲，无论是机械团结，还是有机团结，都体现出社会行动者之间的凝聚性。在齐美尔看来，二方关系和三方关系等小规模群体中才更能体现社会团结的含义，例如三方关系中大量存在分而治之、联盟等现象。当然，对于社会理论家来说，社会团结是一种集体层面的意识问题，这涉及意识形态问题。对于结构主义社会学家来说，他更加关注的是网络的结构性质如何带来团结和凝聚力。

当我们说"这个群体真有凝聚力"的时候，可以有多种含义。例如，群体成员之间的联络很多，能够很高兴地"聚在一起"，众志成城，有一种"力量"使得人们保持一心，能够抵抗外部的破坏性力量等。然而，这些说法多数没有区分作为"社会事实"的"团结"与基于"团结"而导致的心理层面或行为层面的"结果"。

本章探讨凝聚力和关联性的含义、关联性与小世界之间的关系，并且对此进行测量。本章内容主要参考了美国杜克大学社会学系詹姆斯·穆迪（James Moody）教授个人主页上的讲义，网络地址参见穆迪（Moody，2008）。

第一节　关联度和凝聚力

关联度的定义可以有多种。可以将关联度定义如下：关联度指的是一个集

体的成员相互联络的程度。关于什么是凝聚力，也没有一个公认的答案。本节给出其描述性定义和操作性定义。

一、根据社会关系界定的凝聚力

凝聚力指的是一个集体的全部成员通过社会关系联系在一起的程度，这是凝聚力的描述性定义。关于关联度和凝聚力，可以进一步解说如下。“关联度”主要是针对群体中行动者的联络程度而给出的一个概念，可以说某个点的关联度的大小，也可以谈整个网络的关联度大还是小，不过关联度更关注“关联程度”。凝聚度则是针对整个网络来谈的，因而可以说某个网络的凝聚力的大小，但一般不说某个点有凝聚力。凝聚力的这个定义体现了其五维要素：它关注的是什么因素将群体凝聚在一起，表达了凝聚力的群体性质，这个概念可体现凝聚力的连续性，依赖于可观察的社会关系，可应用于任何规模的群体。凝聚力的这个定义体现了它与关联度的关系。

首先，要想网络有“凝聚力”，其中的行动者必须是“关联”的，否则该集体不可能有凝聚力。不过，即便行动者都是“关联”的，该集体也未必有凝聚力。也就是说，关联度是凝聚力的必要不充分条件。

其次，网络要想有“凝聚力”，其各个点必须具有可达性(reachability)，即任何两点之间都至少存在一条途径。两个点之间的途径越多，点的关联度就越大，但是整个网络的凝聚力未必越大。如果整个网络中的任何两点之间都有较多途径，那么整个网络的凝聚力就越大。换句话说，关系的密度往往不是决定性的因素，关系的模式(pattern)才更重要。例如，下面两个图的密度相同，但是由于关联的模式不同，可达性却截然相反。一个图是可达的连通图；另外一个图却是不连通图，分为两个成分。这两个图中的点都有关联度，但是右图的凝聚力明显小，因为它分成了两个子网络。

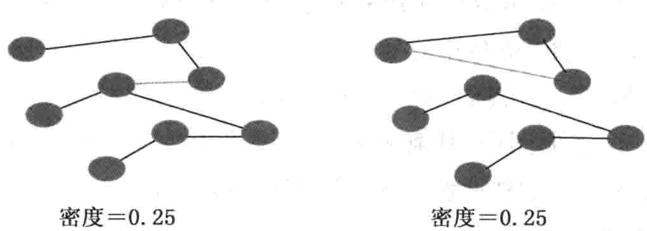
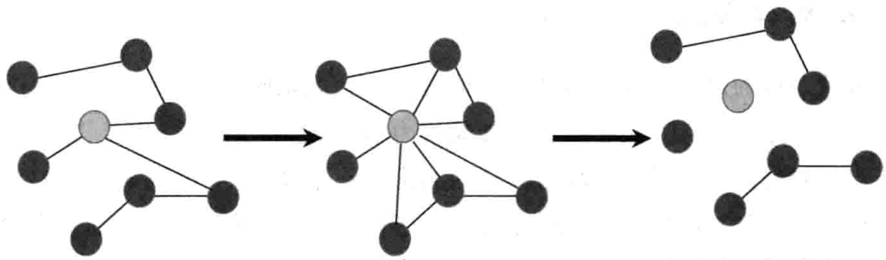


图 7.1　联通图和不联通图

如果一个网络的密度增加，不过是通过一个核心人物的努力而增加的，会有什么效果？请看如下三个图。



密度=0.25　　密度=0.39　　取走一点，整个网络就分崩离析

图 7.2　依赖于一个点的联通图

由图 7.2 可见，对于规模相同的两个网络来说，如果密度只是通过一个核心点的努力而增加的，那么密度大的网络将对核心点产生很大的依赖，因而是“不稳健”的。这就涉及“独立途径”这个概念，需要介绍凝聚力的第二种定义。

二、根据独立途径数界定的凝聚力

对于一个群体的任何一对成员来讲，如果其中都存在多个独立途径将二者联系在一起，我们就说该群体在结构上具有凝聚力。不过这里要介绍什么是“独立途径”。独立途径指的是除了起点和终点一样外，其余点都不同的途径。

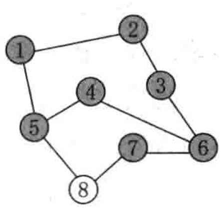


图 7.3　8 点图

比如，对于上面这个图来说，在点 1 和点 8 之间存在着多条途径，如 1-5-8、1-5-4-6-7-8、1-2-3-6-7-8、1-2-3-6-4-5-8，但是只有两条途径是独立的，前两条不独立，因为它们都要经历点 5，后两条也不独立，因为它们都要经历点 2、点 3、点 6。

对于规模和密度都相同的两个联通图来说，如果其中一个图的很多线都通过一个人，那么该图将具有较小的凝聚力(如图 7.4 的左图所示)；反之，如果一个图中的线不是围绕着一个点展开的，那么该图将具有较大的凝聚力(如图 7.4 的右图所示)。在一个连通图中，其凝聚力随着网络中独立途径数目(independent paths)的增加而增加。这种见解应该不会让读者感到困惑。试想，如果一个单位的成员多数“围绕领导转”，使得少数几个人的度数很高，这能是一个凝聚力大的

单位吗？反之，如果一个单位的成员相互之间都有一些联系，并不“围着领导转”，这恰恰表明该单位是一个“团结”的、“有战斗力”的集体！因为它不会因为某位“领导”的退场而分崩离析。

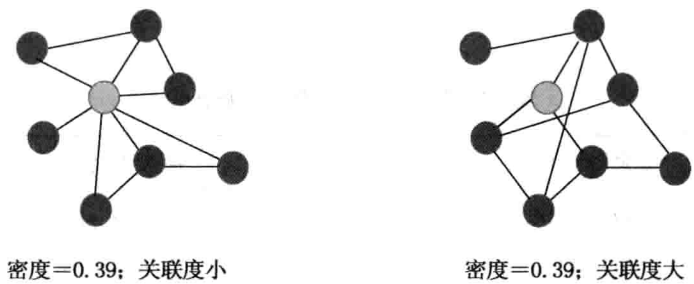


图 7.4　密度相同但关联度不同的两个图

请看下面四个图。图(1)分成两个子图，两个子图之间不存在独立的途径，其凝聚力很小。图(2)将两个子图联络在一起，不过只通过一个点 7 连在一起，独立的途径数是 k＝1，整个图的凝聚力大了一些。图(3)通过点 6 和点 13 连在一起，独立的途径数是 k＝2。图(4)通过点 6、点 7、点 9 所在的途径连在一起，独立的途径数是 k＝3，该网络的凝聚力最大。

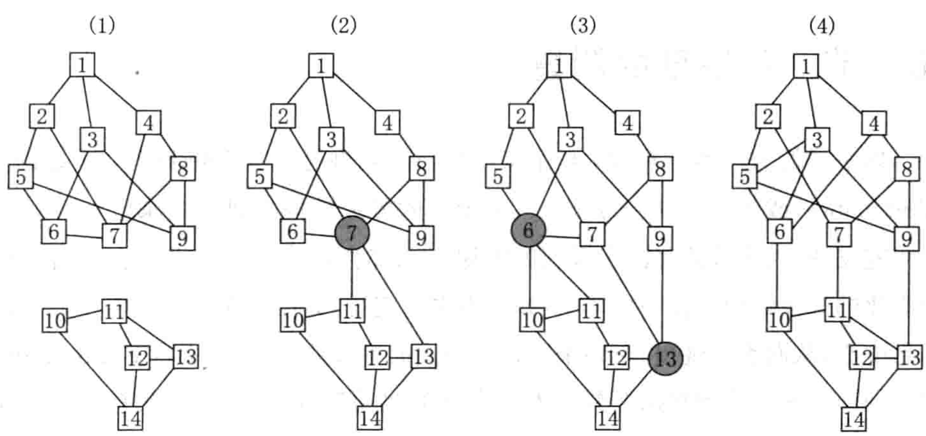


图 7.5　关联水平不同的四个图

基于上述讲解，我们可以给出凝聚力的操作化定义。对于一个群体来说，其中的每一对行动者之间都存在独立途径数，这些独立途径数中的最小者就

是该群体的结构凝聚度。这个定义等价于下面的陈述。对于一个连通的群体来说,为了使其变得不连通需要去掉一些点,所去掉的最小点数就是该群体的凝聚力。

总之,凝聚力高的图和凝聚力低的图在很多方面有所不同,二者之间的差异如表 7.1 所示。

表 7.1 凝聚力不同的图的性质比较

凝聚力低的图	凝聚力高的图
权力集中	权力分散
信息集中	信息分散
行动者不平等	行动者平等
易受到个别点的影响	不易受个别点的影响
分派结构	均匀结构

关联度和凝聚力研究具有很多理论意义。例如,结构上具有凝聚力的群体能更快地传递信息,一个社群就是在结构上有凝聚力的群体,通过资源/信息的流动,可以在一个具有结构凝聚力的群体内部实现再生产。结构凝聚力高的群体成员之间更加平等,因为没有人会控制资源。

第二节 关联度的测量

如果我们关注有向网络,就存在一个与关联性息息相关的概念,即支配性(dominance)或等级性(hierarchy)。关于图的等级结构的研究由来已久。有学者研究发现,大多数拥有民主倾向的志愿组织都会随着时间的推移表现出一种中心化的等级结构。在小群体中,如果不限定交往的对象,那么小群体会集中到信息中心,从而也表现出"车辐条"(wheel)式的等级结构来(转引自 Krackhardt,1994:89)。著名管理学家西蒙等人也提过类似的观点,但是一直没有学者研究如何测量等级性,这就是克拉克哈特(1994)的任务。

为了方便研究图的等级结构性,有必要首先给出一个公认的等级结构,用它作为比较的标准。克拉克哈特认为,如图 7.6 所示的单向树形结构(outtree)就是等级性的原型(archetype),它是一种标准结构。

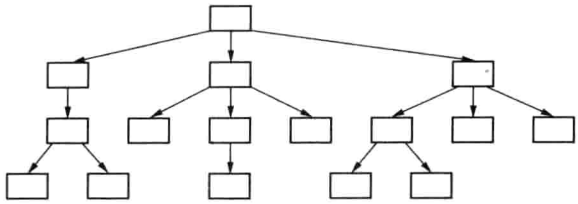


图 7.6　一个标准的树形图

需要强调单向树形结构的以下特点。第一，它是一个有向图；第二，除了“最高点”以外，其他点得到的箭头数只有一个，即只得到一种关系，但是它发出的关系可以有多个。如果这种关系代表权威关系的话，那么第二个特点意味着每个点有且只有一个“老板”，但是每个点都可以拥有多个“下属”。

上述说明只是在直观上对单向树形结构的语言描述。如果用图论语言，则可以认为具有如下特点的有向图就是单向树形结构：

1. 有向图有关联性(connectedness)；
2. 有向图拥有等级结构(digraph hierarchy)；
3. 有向图有效率(digraph efficiency)；
4. 有向图中每一对点都拥有一个最低上限(least upper bound)。

下面介绍这四种维度的测量(Krackhardt，1994)(本部分参见穆迪教授个人主页上的 ppt 讲义)，即有向图的关联度、等级度、效率和最近上限。

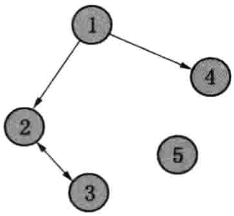
一、关联度

(一) 网络的关联度

对于一个有向图来说，如果其中的任何两点之间都可以建立联系，则称这样的图为关联图，关联图也叫做成分。问题是，对于一个网络来说，其关联的程度有多大？如何测量？我们可以通过“可达性”来测量关联程度。

可以想象，如果一个图中有一些点相互之间不可达，那么这样的图的关联性一定较小。由此可以给出关联度的测量公式：$C=1-\left[\frac{V}{N(N-1)/2}\right]$，其中 V 是该网络中不可达的点对数目，N 是网络的规模。

从操作化的角度讲，“关联性”的计算需要用到可达矩阵，下面用一例加以说明。



有向矩阵

	1	2	3	4	5
1	0	1	0	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

无向矩阵

	1	2	3	4	5
1	0	1	0	1	0
2	1	0	1	0	0
3	0	1	0	0	0
4	1	0	0	0	0
5	0	0	0	0	0

可达距离矩阵

	1	2	3	4	5
1	0	1	2	1	0
2	1	0	1	2	0
3	2	1	0	3	0
4	1	2	3	0	0
5	0	0	0	0	0

图 7.7　一个 5 点图及其矩阵

对于上述 5 点图来说，如何计算其关联度？首先，构造与之对应的有向矩阵；然后，对该矩阵进行对称化处理(在 UCINET 中，沿着 Transform→Symmetrize，选择上述有向矩阵，即可进行对称化处理)，得到无向矩阵；最后，在 UCINET 中，利用 Network→Cohesion→Distance 路径，选择刚刚得到的无向矩阵进行分析，即可计算出可达距离矩阵。

V＝可达矩阵对角线上方"0"的总数①。所以，从可达矩阵中可以看出 $V=4$，$N=5$。所以，$C=1-\left[\dfrac{V}{N(N-1)/2}\right]=1-\left[\dfrac{4}{5(5-1)/2}\right]=0.6$。

实际上，该值恰好等于如下可达矩阵(reachability matrix)的密度。在 UCINET 中，利用 Network→Cohesion→Reachability 路径，选择上述无向矩阵，即可计算出如表 7.2 所示的可达矩阵(该矩阵只表明两点之间是否可达，不关注可达的距离)。

表 7.2　可达矩阵

	1	2	3	4	5
1	0	1	1	1	0
2	1	0	1	1	0
3	1	1	0	1	0
4	1	1	1	0	0
5	0	0	0	0	0

该矩阵的密度＝(矩阵中 1 的个数)/$(N(N-1))=12/(5\times4)=0.6$(或者利用 UCINET 中的 Network→Cohesion→Density 路径计算)。可见，可达矩阵的密度就等于关联度。

---

① 如果网络规模很大，我们不能用肉眼数出来矩阵中有多少个"0"，这时候需要用到 UCINET 中的矩阵代数算法来计算，具体算法可参见 UCINET 中的 Matrix Algebra。

（二）网络中某点的关联度

上述关联度思想关注的是整个网络的关联程度。我们也可以探讨网络中某个点的关联性。对于一个网络中的某个与其他点相连的点来说，如果去掉与之相连的一些点，那么该点就可能不能达到其他点，也就没有关联度。问题是，去掉多少个点才能使该点不可达到其他点？在 UCINET 中，可以利用 Network→Cohesion→Point Connectivity 程序计算出来，该程序会计算出上述去掉的点数。如果连接两个点之间的路径有多条，那么这两个点就具有高度的"关联性"，表 7.3 表示的是利用上述程序对诺克（David Knoke）教授搜集的 10 个组织之间信息传递网络（从 UCINET 自带的 KNOKBUR 数据中打开，其中一个就是信息网络 KNOKI）中的点关联度（point connectivity）分析结果。

表 7.3　诺克信息传递网络中的点关联度

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	4	2	3	4	1	4	2	3	2
2	5	0	2	5	7	1	7	2	5	2
3	4	6	0	4	6	1	6	2	4	2
4	3	4	2	0	4	1	4	2	3	2
5	5	8	3	5	0	1	8	2	5	2
6	3	3	3	3	3	0	3	2	3	2
7	3	3	2	3	3	1	0	2	3	2
8	4	6	2	4	6	1	6	0	4	2
9	3	3	2	3	3	1	3	2	0	2
10	4	5	3	4	5	1	5	2	4	0

该结果证明，只要去掉一个关系，第六个组织就失去了与其他成员之间的联系，信息将不能传递！因此，它既是信息的发出者，也是信息的接受者。如果研究行动者之间的独立性或者脆弱性（vulnerability），那么点关联度分析程序就是一个有用的工具。

二、图的等级度

在社会网络分析领域，如何计量支配性或等级性？克拉克哈特指出，图的等级度（Graph Hierarchy，GH）表达的是人们相互之间在多大程度上非对称地可达（asymmetrically reachable）。

图的等级度的计算公式为 $GH=1-\frac{V}{\max(V)}$，其中 V 等于网络中对称地可达的点对数（the number of symmetrically reachable pairs）。$\max(V)$ 等于 i 可

达 j 或者 j 可达 i 的点对数目。显然,图的等级度往往针对有向图而言的。对于上例来说,其有向矩阵、可达距离矩阵和二值化可达矩阵如下:

表 7.4 5 点图的矩阵、可达矩阵及其二值化

有向矩阵

	1	2	3	4	5
1	0	1	0	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

D_{reach}

	1	2	3	4	5
1	0	1	2	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

二值化的可达矩阵

	1	2	3	4	5
1	0	1	1	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

显然,对称地可达的点对只有一对,即 $V=1$;$\max(V)=4$;$GH=1-\dfrac{V}{\max(V)}=1-\dfrac{1}{4}=0.75$。

由上述分析可以看出 GH 的含义,即 GH 越大,表明网络越具有等级结构;反之亦反。

三、图的效率

图的效率(Graph Efficiency, GE)指的是在已知图中所包含的成分数确定的情况下,图在多大程度上存在着多余的线。

图的效率的计算公式为 $GE=1-\dfrac{V}{\max(V)}$。其中,V 是多余线的条数,max(V)是最大可能的多余线的条数(the maximum possible number of excess links)。该公式与图的等级度的计算公式类似,但是含义不同。下面举例加以说明。

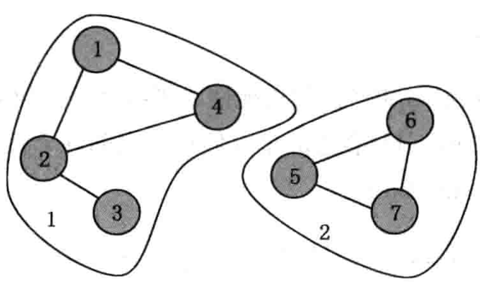


图 7.8 二成分图

在一个规模为 N 的关联成分(connected component)中,从理论上讲它包含的线的条数最小为 $N-1$(假设成分是对称的,只需要利用邻接矩阵上半部分即可)。

图 7.8 包含两个成分。第一个成分包含 4 个点,因此,线的最小数目为 3;而实际上其中共存在 4 条线,因此,多余了 1 条线,即 $V_1=4-3=1$。第二个成分包含 3 个点,线的最小数目为 2;其中实际存在 3 条线,因此,$V_2=3-2=1$。就全部成分来说,$V=V_1+V_2=2$。

如果图中任何两点之间都相连的话(在这种情况下该图就是一个派系了),那么多余线的条数就达到最大值 $N(N-1)/2-(N-1)$。对于第一个成分来说,$\max(V_1)=(6-3)=3$;对于第二个成分来说,$\max(V_2)=(3-2)=1$,因此,$\max(V)=\max(V_1)+\max(V_2)=4$。因此,图的效率 $GE=1-\frac{V}{\max(V)}=1-2/4=0.5$。

对于一般的无向图来说,计算其效率的步骤如下:

1. 区分出图中的全部成分(在 UCINET 中,沿着 Network→Regions→Components 进行成分分析。通常情况下,网络都是关联的,因而只有一个成分);

2. 对于每个成分 i 来说,需要做两步计算,即计算 V(多余线的条数)和 $\max(V)$(最大可能的多余线的条数);具体来讲,分以下两步:

(1) 计算多余线的条数 $V_i=Sum(G_i)-(N_i-1)$(在 UCINET 中,沿着 Tools→Univariate Stats,然后选择需要分析的数据,即可计算出该矩阵的一些统计指标,包括包含的线数);

(2) 计算最大可能的多余线的条数 $\max(V_i)=N_i(N_i-1)/2-(N_i-1)$;

3. 把根据每个成分计算出来的 V 进行汇总,同时也对根据每个成分计算出来的 max(V)进行加总,即 $V=S(V_i)$, $\max(V)=S(\max(V_i))$;

4. 计算图的效率 $GE=1-\frac{V}{\max(V)}$。

实际上,该值必然等于图中各个成分的密度的平均值。

四、最近上限

最近上限(Least Upper Boundedness, LUB)这个值关注的是树形图中存在

多少个“根”(roots)。对于任何一对行动者来说,LUB 就是能够达到二者的最接近的人。在一个正式等级结构中,一对行动者在多数情况下都至少拥有一个 LUB。

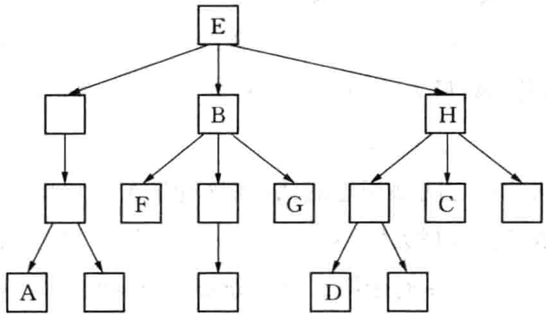


图 7.9 含有一个根的等级树形图

在图 7.9 中,E 是(A, D)的 LUB, B 是(F, G)的 LUB, H 是(D, C)的 LUB。如果在组织中两个人之间没有共同的老板的话,那么根据定义,LUB 是不存在的。例如,在图 7.10 中,4 和 7 就没有 LUB。

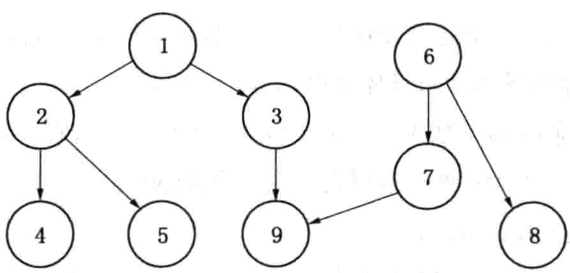


图 7.10 含有两个根的等级树形图

LUB 的计算公式为 $LUB=1-\frac{V}{\max(V)}$。其中分子 V 等于网络中不存在 LUB 的点对数(number of pairs that have no LUB),这个值的计算要根据全部成分进行汇总;分母 $\max(V)=\frac{(N_n-1)(N_n-2)}{2}$。LUB 越大,表明等级性越大;反之亦反。LUB 的具体计算要用到可达矩阵。

以图 7.10 为例,具体计算步骤示例如下。

表 7.5　9 点图的距离矩阵、可达矩阵及其转置

距离矩阵

	1	2	3	4	5	6	7	8	9
1	1	1	1	2	2				2
2		1		1	1				
3			1						1
4				1					
5					1				
6						1	1	1	2
7							1		1
8								1	
9									1

可达矩阵 R

	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1				1
2		1		1	1				
3			1						1
4				1					
5					1				
6						1	1	1	1
7							1		1
8								1	
9									1

可达矩阵的转置 R′

	1	2	3	4	5	6	7	8	9
1	1								
2	1	1							
3	1		1						
4	1	1		1					
5	1	1			1				
6						1			
7						1	1		
8						1		1	
9	1		1			1	1		1

注意，上述矩阵对角线的值被设定为 1，这是因为我们假定，如果一对点不能被至少一个共同点达到，就违反了 LUB 的定义。我们可以通过如下所示的矩阵操作获得“共同的可达性”(common reachability)。

下面计算有多少个点能够“共同可达到”某两个点。对于两个点(如 A 和 B)中的任意一个点(如 A)来说，我们需要首先找到可达到 A 的点(如 C)，再看 C 点是否可达到 B，如果可达，说明 C 是 A、B 的共同可达的点。按照这种思路，我们需要先计算可达矩阵的转置矩阵(因为该矩阵的行表达的就是接受的关系)，再乘以可达矩阵(其行表示发出的关系)，得到的就是“共同可达的矩阵”。如果该矩阵中的值为 1，即表明 1 之所在的行和列之间是共同可达的。也就是说，在二者之间存在着共同可达的点。总之，用“可达矩阵的转置矩阵”乘以“可达矩阵”，得到的就是“共同可达的矩阵”。用矩阵语言表示就是：$R' * R = CR$。

表 7.6　9 点图的共同可达矩阵

共同可达矩阵＝R′*R＝CR＝

	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	0	0	0	1
2	1	2	1	2	2	0	0	0	1
3	1	1	2	1	1	0	0	0	2
4	1	2	1	3	2	0	0	0	1
5	1	2	1	2	3	0	0	0	1
6	0	0	0	0	0	1	1	1	1
7	0	0	0	0	0	1	2	1	2
8	0	0	0	0	0	1	1	2	1
9	1	1	2	1	1	1	2	1	5

在共同可达矩阵中，任何 0 值都表明对应的一对点之间不存在 LUB。根据

计算公式 $LUB=1-\frac{V}{\max(V)}$，首先计算 V。由于 V 等于网络中不存在 LUB 的点的对数，就本例而言，$V=15$（用肉眼即可看出"共同可达矩阵上半部分有 15 个 0"）。又由于 $\max(V_n)=\frac{(N_n-1)(N_n-2)}{2}=\frac{(9-1)(9-2)}{2}=28$。因此，该图的"最近上限"指标数为 $LUB=1-\frac{V}{\max(V)}=1-\frac{15}{28}=0.4643$。

当然，针对具体的复杂网络关系来讲，我们不能用手工计算。上述指数的计算都可以在 UCINET 中的矩阵算法中进行。也可以在 UCINET 中，点击 Network→Cohesion→Krackhardt GTD 即可直接计算出这些指数。例如，就上述 9 点图来说，选择与上图对应的数据矩阵（如表 7.7 所示），点击"Krackhardt GTD"后，即可计算出这四类指数。

表 7.7　9 点图的多种关联指标

	1	2	3	4	5	6	7	8	9
1	0	1	1	0	0	0	0	0	0
2	0	0	0	1	1	0	0	0	0
3	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	1	0
7	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

→

Krackhardt GTD Measures

	1
Connectedness	1.0000
Hierarchy	1.0000
Efficiency	1.0000
LUB	0.4643

可见，该网络的关联性、等级性和效率指标都是 1，而 LUB 指标则是0.4643，与前面的计算一致。

在一个非正式网络中，LUB 保证了组织内部差异或者冲突得以解决。对于一对行动者来说，如果它们之间存在一个 LUB，这个 LUB 人就有一种解决二者之间冲突的潜力。如果拥有 LUB 的行动者较少，可以想象冲突就难以解决（Krackhardt，1994：100）。

最后需要指出的是，关于等级性的研究是多种多样的。本章只探讨了其中的一些算法，除此之外，中心性研究实际上也是等级研究的一类，本书对此已有论述。

第三节　关联性与小世界

有关关联性的研究还与当代著名的"小世界"（small world）研究息息相关，

因为小世界就是一个关联网。在小世界研究中，最初关注的问题是，世界上任何两个人能够相互认识的概率是多少？把该问题进行推广，人们之间通过 1 步，2 步，3 步，…，k 步建立联系的概率是多少？与之相关的问题是，如果人们能够建立联系，那么他们之间的疾病也会相互传染，因此，我们可以利用模型对疾病的传播、信息的传递等进行模拟。有关这种可达性的结构（reachability structure），可通过比较网络中的途径与随机网络中的途径得以揭示。

一、小世界的含义

崔弗斯和米尔葛伦（Travers & Milgram，1969）通过小群体实验，得到一个著名论断，“世界上任何人之间都通过大约 6 步就可以建立联系”，因此，整个世界是小世界（small world）。这里有一些问题值得讨论，假设我们知道了网络的存在，那么把任何两个人分开的最长的途径（用可以握手的关系来界定）是多少？6 步距离是大还是小？人们在传递信息的时候是如何考虑选择哪些人的？

可以想象，两个人之间完全可能通过 10 多个步骤才建立联系，也可能通过很少的步骤联系在一起。例如，在理想情况下，假设两个人都认识 150 个人，那么全世界的人经过 4 个中间人就可以建立联系了。现在的问题是，如果人们之间是随机认识的，那么根据随机网络理论，世界上的人相互认识的速度相当快。

例如，假设有一个随机的密友网络。如果根据密友的数量来研究可达性，那么随着密友数量的增加，该随机网络的可达性的增速更快。如果平均每个人的密友数量为 2 个，那么在经过大约 15 步之后，每个人接触的人数将达到整个网络人数的 80%。如果平均每个人的密友数量为 3，那么在经过大约 8 步之后，行动者可达的数目将超过整个网络的 90%。如果平均每个人的密友数量为 4 个，那么在经过 6 步之后，就可达到接近整个网络的 97%。如图 7.11 所示（Moody，2008：Class 6）。

在崔弗斯和米尔葛伦（Travers & Milgram，1969）之后 30 年，沃茨（Watts，1999）继续此类话题，他追问的问题是：为什么我们看到了小世界现象？它对于社会系统的动态性质有何意义？沃茨的贡献在于，他明确指出总体上的重大变化可能来自局部不显著的网络变动。

沃茨指出，小世界之所以令人感兴趣，原因在于它拥有如下四个方面的性质（Watts，1999：495—496）：

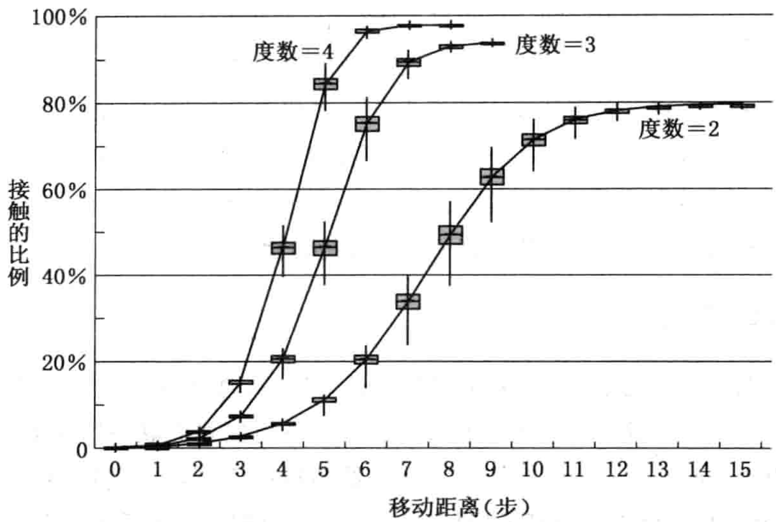


图 7.11 随机密友网络中密友数量与可达性之间的关系

1. 整个网络巨大,现实世界中包含的人数达到十亿数量级;

2. 网络是稀疏的。平均来说,人们在日常生活中接触到的人数极少,最多与几千人接触,而即便几千人相对于几十亿来说也微乎其微,因此网络是稀疏的;

3. 网络是去中心化的,即不存在核心点;

4. 整个网络是高度聚类的,大多数朋友圈都有重叠。

前两个条件在现实生活中是能够满足的,后两个条件虽然比较难以保证,但在现实生活中是可以理解的。满足这些条件的网络就是小世界。

二、小世界的测量指标

上述关于小世界的说明都是定性的,为了精确起见,我们可以对小世界现象进行形式化表征。对于关联网络来说,我们可以利用如下两个统计量来刻画其性质。

1. 特征途径长度(characteristic path length)L:连接任何两个点之间最短途径的平均长度(这一点只适用于关联网络)。L 是测量网络整体性质的一种测度。

2. 聚类系数(clustering coefficient)C。与 L 不同的是,聚类系数 C 是一种关于局部网络结构的指标,该指标有两种界定方式:

第一，平均局部密度(average local density)。如果用 C_v 表示点 v 的聚类系数，那么 C_v＝v 点的个体网密度。整个网络的聚类系数等于各个点密度系数的均值，即 $C=C_v/n$。

第二，传递性比率(transitivity ratio)：封闭三方组的数量(number of closed triads)与三方组总量之比。

所谓小世界图(small world graph)就是任何一个拥有相对较小的 L 和相对较大 C 的图。问题是，什么样的指标算做"较大"或者"较小"？沃茨给出了判断 L 和 C 大小的三个标准(Watts，1999:499)：

1. 总体规模 n 是固定的；

2. 各个点的平均度数 k 也是固定的，k 远远大于 1 并且远远小于 n，这样的话，网络就是稀疏的，但是有一定的密度，从而保证网络有其可能的结构；

3. 网络必须是关联的，即任何两个点都可以通过一定的长度建立与其他点的联系。

沃茨(Watts，1999:500—501)构造了一种聚类系数最高的图，称之为"洞穴人图"(Caveman graph)，如图 7.12 所示。

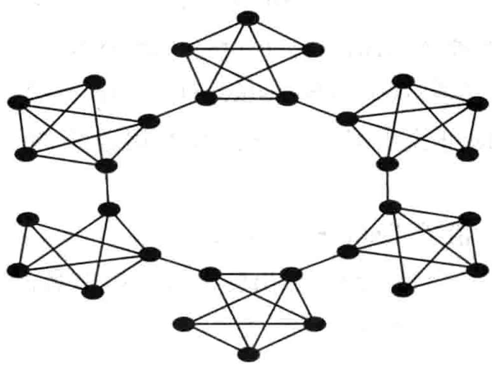


图 7.12　洞穴人图

该图的聚类系数为 $C_{caveman} \approx 1-\frac{6}{k^2-1}$，L 系数为 $L_{caveman} \approx \frac{n}{2k+1}$

沃茨证明了，在一个高度聚类的有序网络中，单个随机的联系就会产生一条捷径(shortcut)，从而大大降低了 L 的值，进而证明，如果一个图拥有少量的捷径，该图就变成小世界图。无论是互联网、电影明星网，还是电网，都表现出小世界的特征。

这个结论的重要意义还在于对疾病传播的研究上。由上述论述可知,在一个高度聚类的网络中疾病的传播速度很慢。但是,疾病传递的动态机制是非线性的,并不表现出明显的模式来,随着局部的微小调动(捷径),整个网络的结构会出现巨大变化(如疾病迅速传播,信息迅速传递等)。

三、小世界的测量示例

(一) 聚类系数的计算

在 UCINET 中,要计算一个网络的聚类系数,首先要导入该网络,对网络进行对称化处理(Transform→Symmetrize→Maximum),然后沿着 Network→Cohesion→Clustering Coefficient 这条指令计算聚类系数。注意,该程序会计算出两类聚类系数。一是根据局部密度计算出来的聚类系数,即 overall graph clustering coefficient。二是根据传递性计算出来的聚类系数,即 weighted overall graph clustering coefficient。该程序同时给出每个点的聚类系数计算结果,具体例子略。

(二) 途径长度均值的计算示例

在 UCINET 中,要计算一个网络的途径长度均值,首先也要导入该网络,对数据进行对称化处理,然后沿着 Network→Cohesion→Distance 计算距离,得到"距离矩阵",最后沿着 Tools→Statistics→Univariate→Matrix,对得到的"距离矩阵"进行计算,即可算出 C 值。下面我们用假想的一个随机数据加以展示。

在 UCINET 中,沿着 Data→Random→Sociometric 路径构造一个拥有 3 000 个点的社群矩阵,令每个点的点出度都为 150,命名该数据为 Sociometric Random Graph,如图 7.13 所示。

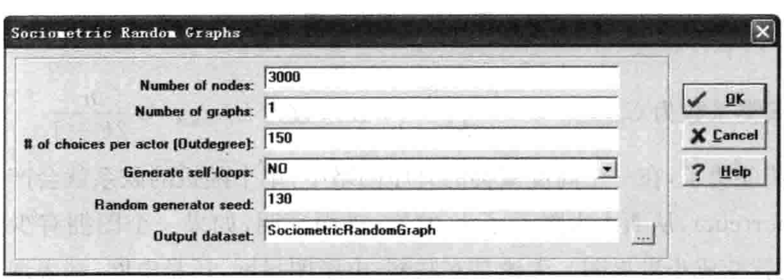


图 7.13 随机图的建构对话框

然后沿着 Network→Cohesion→Distance 即可计算出 Sociometric Random Graph 数据的距离矩阵 Geodesic Distance，该矩阵的信息如下。

表 7.8　规模为 3 000 的随机图距离信息

```
GEODESIC DISTANCE
------------------------------------------------------------------------------
Type of data: ADJACENCY
For each pair of nodes, the algorithm finds the # of edges in the shortest path between
them.
Average distance                                   = 1.950
Distance-based cohesion ("Compactness")            = 0.525
(range 0 to 1; larger values indicate greater cohesiveness)
Distance-weighted fragmentation ("Breadth")        = 0.475

Frequencies of Geodesic Distances
                              Frequency        Proportion
                           ------------      ------------
                       1    450 000.000             0.050
                       2  8 543 040.000             0.950
                       3      3 960.000             0.000
                Distance matrix saved as dataset GeodesicDistance
```

由这个计算结果可以看出，距离是 1 的情况出现了 450 000 次，距离是 2 的情况出现了 8 543 040 次，占总数 8 997 000(＝3 000×2 999)的 95％。这已经说明了绝大多数人之间的距离是 2。

我们也可以对距离进行描述统计分析。沿着 Tools→Statistics→Univariate Stats，点击后出现一个对话框。在对话框的“Input dataset”中选择上述计算中程序直接产生的距离矩阵 GeodesicDistance，在“Which dimension to analyse”框中选择 Matrices，点击 OK 后即可计算出距离的各项指标，如表 7.9 所示。

表 7.9　规模为 3 000 的随机图的统计量

```
         UNIVARIATE STATISTICS
-----------------------------------
                                  1
                     --------------
1         Mean                1.950
2      Std Dev                0.219
3          Sum       17 547 960.000
4     Variance                0.048
5          SSQ       34 657 800.000
6        MCSSQ          431 846.875
```

7	Euc Norm	5 887.087
8	Minimum	1.000
9	Maximum	3.000
10	N of Obs	8 997 000.000

可见,在这个 3 000×3 000 距离矩阵中,平均距离为 1.95,标准差是 0.219,方差是 0.048,最小距离是 1,最大值是 3。也就是说,在由这 3 000 个人构成的较大网络中,按照每个人认识 150 人计算的话,那么这个关联网络中的任何两个人之间的距离平均值仅仅是 1.95,最大距离不过是 3,况且距离为 3 的情况极少。两个人的距离是 2,这意味着两个人之间仅仅有一个中间人。可见,在该网络中,人们之间平均只需要通过一个中间人就可以建立联系了。

本章小结

只要一个网络的成员之间都可达,就称这样的网络具有关联性。关联性随着网络中独立途径数目的增加而增加,因此,关联的模式比较重要。关联性的指标包括关联度、等级结构、效率和最低上限。关联性与小世界息息相关。如果一个网络巨大,其中的关系稀疏,不存在核心但是高度聚类,这样的网络就叫做小世界,其测量指标包括特征途径长度和聚类系数。

第八章

块模型分析

根据劳瑞和怀特(Lorreian & White)的说法,要想找出一个网络的总体模式,必须根据行动者的结构对等性做一些转换性的工作,所使用的方法主要是把各个点集中到更大的点集之中,这样,各个点集系统的内在结构要比在大量孤立点之间的具体关系更加明晰。怀特等学者(White, Boorman & Brieger, 1976)的块模型思想是把社会生活看成是相互关联的角色系统(interconnected system of roles)。他们认为角色在一个角色系统中相互关联,这就是社会结构(social structure)。

第一节　块模型的含义

怀特等学者(White, Boorman & Breiger, 1976)描述了块模型(block-modeling)的理论和经验要素。他们指出,可以根据角色之间的互动(而不仅仅根据帕森斯的宏大理论)来解释社会结构。劳瑞和怀特(Lorreian & White, 1971)又进一步把一个复杂网络"简化"为"块模型"或者"像矩阵"(Image matrix)。在进行"简化"工作的时候,要把初始发生矩阵中的点用一种聚类分析的方法进行重排,从而形成了在结构上对等的一系列像矩阵。也就是说,在社会网络分析中,我们可以根据"结构对等性"对行动者进行分类(参见刘军,2004:第七章),对此进行研究的方法就是块模型分析方法。块模型方法不同于个体网络分析,因为它关注的是网络的总体结构。

例如,在图 8.1 中,集合 M_1 包含一些点,这些点相互之间被看成是在结构上对等的、但是在结构上不同于另一个结构上对等的点集 M_2。劳瑞和怀特认为,一个网络的最基本特征在各个点集之间的关系中可以明显看到,并且这些关系的本质可以通过像矩阵的各个元素(块)体现出来。怀特的大多数后续工作就是研究如何产生这种块模型。

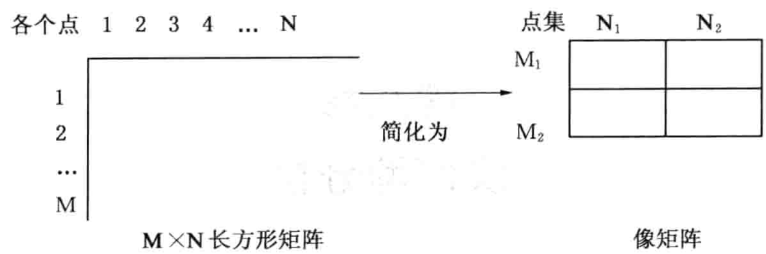


图 8.1 一个网络及其块模型

在构建块模型的时候,需要用到 UCINET 软件中的 CONCOR 程序(参见 Doreian et al., 2000:12)。块模型方法是根据结构性信息把各个点进行分区的方法。它可以利用下面将介绍的 CONCOR 来分析,也可以利用 UCINET 中的层次聚类方法(hierarchical clustering)进行分析。

一、块模型

块模型(blockmodels)分析最早由怀特、布尔曼和布雷格(White, Boorman & Breiger, 1976)提出,它是一种研究网络位置模型的方法,是对社会角色的描述性代数分析。后来,学者们从许多方面对此概念进行了深入研究和推广。随后出现了随机块模型(stochastic blockmodels,参见 Wasserman & Faust, 1994)和一般化的块模型等(Doreian, Batagelj & Ferligoj, 2004, 2005)。除此之外,许多学者还运用"块模型"研究一些具体问题,例如对科学共同体的研究(Breiger, 1976)、对世界经济体系的研究(Snyder & Kick, 1979)、对组织问题的研究以及大量的小群体研究等。总之,"块模型"这个概念和方法已经得到了广泛的应用。

(一) 块模型的定义

块模型的定义可以在三个层次上给出。我们首先给出描述性定义,然后给出形式化定义。

定义 1:一个块模型是由以下两项组成的:(1)把一个网络中的各个行动者按照一定标准分成几个离散的子集,称这些子集为"位置",也可称之为"聚类"、"块"(参见 Wasserman & Faust, 1994:350; White, Boorman & Breiger, 1976:769);(2)考察每个位置之间是否存在关系。根据这种定义可以认为,一个"块"就是邻接矩阵的一部分,是一个整体中的子群体。

这样看来，一个块模型就是一种模型，或者一种关于多元关系网络的假设。它提供的信息是关于各个位置(而不是每个行动者)之间的关系，因而研究的是网络的总体特点(White, Boorman & Breiger, 1976)。

定义 2：一个块模型是把一个网络 N 中的行动者分区成为各个位置 B_1，…，B_2，B_B，并且存在一个对应法则∅，它把行动者分到各个位置之中，即如果行动者 i 处于位置 B_K 之中，则∅(i)＝B_K。我们利用 b_{klr}表征位置 B_K 和 B_l 在关系 X_r 上是否存在联系。如果存在联系，则 $b_{klr}=1$，否则为 0。

定义 3：我们也可以根据像矩阵(image matrix)B＝{b_{klr}}对块模型进行定义。B 是一个 B×B×R 的排列，其要素 b_{klr}的含义与上述相同。整个矩阵 B 也是一个块模型。初始矩阵用常见的 g×g×R 多元关系社群矩阵表示。可见，B 是初始矩阵的简化矩阵。它包含两个成分：对应法则∅(对各个行动者的位置进行指派)和矩阵 B(给出各个位置之间的关系的有无)。每个行动者被指派到并且只指派到一个位置之中。

矩阵 B 中的各项都叫做“块”(block)，每个“块”(即 b_{klr})实际上对应的是初始矩阵的一个子矩阵。如果某块为 1，称之为 1-块；如果为 0，称之为 0-块。

总之，一个块模型就是对一元关系或者多元关系(包括二值关系以及多值关系)网络的一种简化表示，它代表的是该网络的总体结构。每个位置中的各个行动者都具有结构对等性。例如，位置 B_K 中的所有行动者与 B_l 中的所有行动者之间的关系都类似。因此，块模型是在位置层次上的研究，而不是在个体层次上的研究(参见 Wasserman & Faust, 1994:394—340; Burt, 1976:93—122)。

(二) 块模型的构建

学者们提出许多方法构建块模型，这涉及两个步骤。第一步是对行动者进行分区，即把各个行动者分到各个位置之中。常见的方法是 CONCOR 以及层次聚类方法。第二步是根据一些标准确定各个块的取值，即各个块是 1-块，还是 0-块。不同性质的关系采用的标准是不同的。总的来说有六种标准：完全拟合、0-块标准、1-块标准、α-密度指标、最大值标准(适用于多值数据)以及平均值标准(适用于多值数据)(参见 Wasserman & Faust, 1994:397—401)。

前三种标准都是比较严格的，一般情况下用不上。确定 1-块还是 0-块的最常用的标准是 α-密度指标，其中 α 是临界密度值，它通常指的是整个网络的平均密度值。在对多个网络进行分析的时候，α 也可以指各个网络的平均密度值。也就是说，α 可以是一个，也可以是多个。α 还可以用行的平均值来代替。我们

利用整个网络的平均密度值表示 α。

(三)对结果的解释

对块模型分析结果的解释有如下三个层次。

1. 个体层次——利用个体的属性资料(如性别、年龄、社会地位等)分析块模型的有效性,这是因为个体的属性往往与网络的结构关系密切。

2. 位置层次——对各个位置(块)进行描述性分析。要具体考察各个位置之间是如何发送和接收信息的。这种描述性分析有助于总结各个位置发送和接收信息的趋势。

对"子群"之间关系的研究借鉴了"点"之间的研究。在研究网络中行动者的特点时,我们往往利用点入度和点出度将作为行动者的点分为四类:孤立点(isolates);发送点(transmitters),即只有点出度的点;接收点(receivers),即只有点入度的点;传递点(carriers or ordinary points),即既有点入度又有点出度的点(参见 Marsden, 1984, 1990; Wasserman & Faust, 1994:411)。这种分类研究也适用于对网络位置的描述。

伯特(Burt, 1976:93—122)对位置进行了分类研究。他首先区分了两类位置:一类位置的成员接受关系,另一类位置的成员发送关系。其次,他又对如下两类位置进行了区分:一类位置的成员之间的总关系数中有一半以下针对自己位置的成员,另一类位置的成员之间的总关系数中有一半以上针对自己位置的成员。这样,我们就可以确定哪些位置接受关系,哪些位置发送关系,哪些位置内部关系紧密,哪些位置内部关系不紧密。从而产生了与上述类似的四种分类:孤立位置,其成员与外界没有任何联系;谄媚位置(sycophants),其成员与其他位置的成员之间的关系比与自己成员之间的关系多,并且没有接收到多少外来的关系;经纪人位置(brokers),其成员既发送也接受外部关系,而且内部成员之间的联系比较少;首属位置(primary),其成员既接受来自外部成员的关系,也有来自自身成员的关系。

沃瑟曼等人指出,在考察关系的时候,我们也要分析各个位置的规模(参见 Wasserman & Faust, 1994:413)。例如,假设我们分析来自位置 B_K 的各个成员的关系。假设其中有 g_k 个行动者,那么 B_K 内部可能具有的关系总数为 $g_k(g_k-1)$个。在总体中含有 g 个行动者,因此,B_K 位置各个成员的所有可能的关系有 $g_k(g-1)$个。这样,我们可以期待一个位置的总关系的期望比例为 $(g_k(g_k-1))/(g_k(g-1))=(g_k-1)/(g-1)$。我们可以利用这个比例作为评价

位置内部关系趋势的指标。

表 8.1 基于位置内部以及位置之间的关系给出四种位置分类。

表 8.1　四类位置类型

位置内部的关系比例	位置接收到的关系比例	
	≈0	>0
$\geqslant (g_k-1)/(g-1)$	孤立者位置	首属人位置
$\leqslant (g_k-1)/(g-1)$	谄媚人位置	经纪人位置

当然，上述四类位置标签也依赖于关系的内容。如果关系是“负面性的”（如“不喜欢”等），那么首属人位置最好解释为“受蔑视者”、“替罪羊”等。

需要补充的是，对位置的这种分类研究不能进行统计检验。而随机块模型以及单一关系网络的统计分析方法、多元网络分析方法可以提供二人组选择概率的统计分析。

3. 整体层次——利用像矩阵(image matrix)对总体的块进行描述。

最简单的情况是对 2-位置块模型的像矩阵分析。怀特等人（White, Boorman & Breiger, 1976)指出，对于最简单的块模型，即 2-位置块模型来说，其可能存在的状态有 $2^4=16$ 种。由于某些块相互同构，因而实际上有 10 类各不相同的块模型，并且大多数块模型都有明确的解释性意义（参见 Wasserman & Faust, 1994:421)。

下面首先分析最简单的像矩阵——2-位置的像矩阵，然后分析多于 2 个位置的像矩阵，最后简要介绍对多关系网络的像矩阵分析。

(1) 有两个位置的像矩阵。

最简单的像矩阵是包含 2 个位置（简写为 2-位置）的块模型。怀特等人(White, Boorman & Breiger, 1976)指出，对于最简单的块模型，即 2-位置块模型来说，由于有 2 个位置，因而在块模型中有 2×2=4 个元素。又由于每个元素的取值又有两种可能：1-块和 0-块，因此，2-位置所有可能存在的状态共有 $2^4=16$ 种。在这 16 种状态中，由于某些块相同，因此实际上只有 10 类各不相同，即有 10 个同构类，这 16 种像以及 10 个同构类像的具体结构如图 8.2 所示（Wasserman & Faust, 1994:421)。

① 没有任何关系的块，即虚无块：$\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$

② 有一个位置内关系的块：$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$或者$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$

③ 有一个位置间关系的块：$\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$或者$\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}$

④ 有两个位置内关系的块：$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$

⑤ 有两个位置间关系的块：$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$

⑥ 存在两个关系，即自反式并且“向外”关系：$\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}$或者$\begin{bmatrix}0 & 0\\1 & 1\end{bmatrix}$

⑦ 存在两个关系，即自反式并且“向内”关系：$\begin{bmatrix}1 & 0\\1 & 0\end{bmatrix}$或者$\begin{bmatrix}0 & 1\\0 & 1\end{bmatrix}$

⑧ 存在三个关系，其中两个是位置间关系：$\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}$或者$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}$

⑨ 存在三个关系，其中两个是位置内关系：$\begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}$或者$\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$

⑩ 四个关系，即完备关系$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$

图 8.2　2-位置块模型中可能存在的 10 类像矩阵

在上述 10 个同构类中，大多数像都有明确的理论意义。怀特等学者(White, Boorman & Breiger, 1976)对这些像的理论意义进行了说明。例如，像②中有一个凝聚子群，该位置也是孤立者。像③表明一个位置向另外一个位置发送关系，前一种位置扮演“谄媚者”的角色。像④表示的是“纯粹的”自反性结构，如果该关系是“正向”关系，那么④表明的是两个凝聚子群的情况。④还可能表示的是具有内生性质的系统，该系统中的关系仅仅存在于系统内部的子群之内，子群之间不存在关系。④也可能表示在各个孤立地域中的各个国家之间的贸易关系。⑤是“纯”对称关系。如果所研究的关系是“负面”关系的话，那么这种结构反映的将是“对抗”和“敌对”的状态等。⑥也可能表明一种内生系统，其中所有的关系都指向另外一个群体的成员(例如，在一个外婚制系统中，“从另外一个群体中寻找配偶”)。像④(对于正面关系来说)和像⑤

（对于负面关系而言）的结合，产生了与均衡理论一致的结果。根据该理论，处于一个均衡系统中的行动者可以分为不同的聚类，以至于所有的正向"选择"关系是发生在子群内部，所有的负向"选择"关系是发生在子群之间。像⑥根据所作出的"选择"的不同，区分出"积极的"位置和"消极的"位置。像⑦把凝聚子群（像②）的某些方面和差异结构（像③）结合起来，从而表达了一种核心—边缘结构（其中有一个首属者位置和一个谄媚者位置），这种结构也可以被解释为"等级结构"（hierarchy）。像⑧也是一类特殊的结构，怀特等学者（White，Boorman & Breiger，1976）把这种结构描述为一种核心—边缘结构，或者是一种奉承者模式（hanger-on）。另外，像⑧与像⑤类似，其中的关系仅仅存在于各个位置之间。像⑨几乎是一种完备模式，像⑩则是一个完备关系模式，其中任何位置之间没有差异之分。

（2）有两个以上位置的像矩阵。

对于拥有 3 个位置的单种关系模型来说，可能存在的 3×3 矩阵安排有 $2^9=512$ 种，其中共有 104 种同构类。这时候，我们很难做到对所有这些像矩阵进行全面分析。随着位置数目的增加，各不相同的像矩阵（同构类）的数目也急速增加。例如，4-位置模型可能有 2^{16} 种排列矩阵，这时候分析起来更为困难，我们也不可能进行全盘分析。另外，很多像矩阵并没有什么理论意义。出于上述考虑，同时也为了简便起见，我们往往只需要考察一些具有重要理论意义的"理想模型"。下面我们只针对 4-位置的情形，特别考察能够表现出"凝聚子群"、"核心—边缘结构"、"集中趋势"、"等级性"和"传递性"的像矩阵（Wasserman & Faust，1994：423）。

① 凝聚子群。在现实的组织关系系统中大量存在小群体、非正式组织等现象。对于单种多值关系网来说，这种系统所对应的像矩阵中的关系主要是位置之间的关系。这种像矩阵的主对角线上都是 1，即在位置层次上是自反式的关系（尽管在个体层次上可能对"自反式"关系不加以定义）。例如，我们可能对"自己帮助自己"的关系不给出定义，也不加以研究，但是我们可能通过对一个大群体的分析，发现其中存在多个小群体，并且发现每个群体内部的各个成员之间相互帮助，而不同群体成员之间却不存在帮助行为。这种关系模式就是"凝聚子群"关系模式，表现在像矩阵中就是主对角线的值都是 1，其余都是 0。如图 8.3 中的①图所示。

需要注意的是，该块模型中的 1-块并不意味着其中的行动者一定构成一个

图论意义上的派系(clique)。因为这些行动者构成的图不一定是完备子图。

② 核心—边缘结构。核心—边缘结构也是在现实中存在的一种重要社会结构。在这种结构中主要存在两类行动者:核心行动者和边缘行动者。核心行动者相互之间联系紧密,构成凝聚子群;处于边缘地位的行动者相互之间不存在关系或者存在较少的关系,因而不构成凝集子群,但是它们与核心成员之间有关系(参见刘军,2004:第八章)。

在一个块模型中,如果像矩阵中的各个块可以进行置换,使得1-块集中在像矩阵左上半部分,0-块集中在右下半部分(参见图8.3中的②图),那么在该块模型中就明显表现出核心—边缘结构来。在世界政治经济体制中就常常见到这种核心边缘结构。

③ 集中趋势。所谓集中化的结构也不难理解,即在关系网络中所有关系都指向某个位置,或者都是从某个位置中发出来的。这种结构表现在像矩阵中就是,所有的1-块都在同一列(如果关系指向同一个位置)或者在同一行(如果所有关系都是从同一个位置发出来的)中。当然,自反式关系也可能存在于这种结构之中。国外学者在研究社会学专业期刊的引用问题、组织之间关系网络问题的时候,发现都存在这种"集中趋势"结构,参见图8.3中的③图所示。

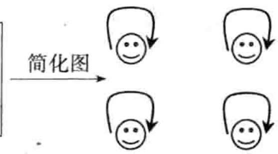
④ 等级性。在军事机构中,军人严格地遵守上级的命令,表现出严格的等级秩序。这种结构表现在像矩阵中就是类似图8.3中的④图的结构。

⑤ 传递性。就三个人A、B、C之间的"帮助关系"来说,如果A帮过B,B帮过C,那么A认为自己也应该帮助C的话,这就是个人层次上的传递性关系结构。在块层次上,关系的传递性是针对"块"(行动者子集或者"位置")而言的。具体地说,就三个位置B_k、B_l和B_m来说,所谓块层次上的传递性关系结构指的是,如果从B_k到B_l存在一个关系,从B_l到B_m也存在一个关系,那么就存在一个从B_k到B_m的关系。

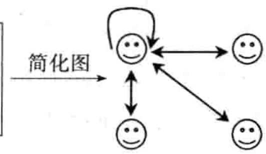
在一个完全传递性的模型之中,我们可以通过一定的矩阵置换步骤,把所有的1-块置换到像矩阵的左下角(或右上角),如图8.3中的⑤所示。

上述各种模型都是理想意义上的模型,在现实世界的关系网络中,像矩阵一般不会表现出上述理想结构来。因此,这些理想的像矩阵类似于韦伯所说的"理想类型"(ideal type),可作为参照物,用来与现实模型进行比较,从而发现现实像矩阵的特点。

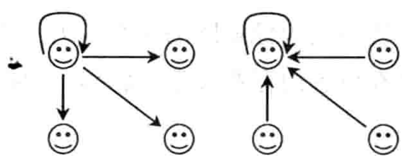
① 凝集子群像矩阵：$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$



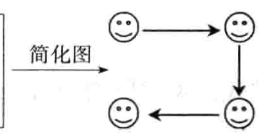
② 核心—边缘像矩阵：$\begin{bmatrix}1&1&1&1\\1&0&0&0\\1&0&0&0\\1&0&0&0\end{bmatrix}$



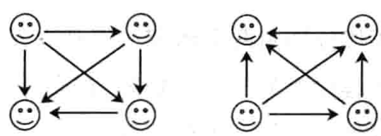
③ 中心化趋势的像矩阵：$\begin{bmatrix}1&1&1&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}$ 或者 $\begin{bmatrix}1&0&0&0\\1&0&0&0\\1&0&0&0\\1&0&0&0\end{bmatrix}$，简化图分别如下：



④ 等级层次的像矩阵：$\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}$



⑤ 关系传递性的像矩阵：$\begin{bmatrix}0&1&1&1\\0&0&1&1\\0&0&0&1\\0&0&0&0\end{bmatrix}$ 或者 $\begin{bmatrix}0&0&0&0\\1&0&0&0\\1&1&0&0\\1&1&1&0\end{bmatrix}$。简化图分别如下：



资料来源：Wasserman & Faust，1994：423。

图 8.3 包含两个以上位置的块模型中的五个理想像

(3) 多元关系的像矩阵。

多元关系的像矩阵比较复杂，难以解释。较好的方法是分别比较各个像矩阵，例如可以分析关系的重叠性（考察像矩阵是否类似），分析关系的交换性、互惠性（如果两个矩阵相互为转置矩阵，则出现关系的互惠性）。在分析的时候应

该结合一定的理论命题。对块模型的操作化研究要利用 CONCOR 方法和程序。

研究多元关系的一种方法是考察各对像矩阵,分析二者是否表现出同样的多元关系模式,例如"交换性"或者"多元性"(multiplexity)。所谓"关系的多元性"指的是两种或者多种关系共同发生的一种趋势。例如,"是朋友"和"是讨论问题的伙伴"这两种关系在两个人之间有可能共同发生。如果两个或者多个像矩阵是恒等(或者几乎恒等)的,该块模型中就明显出现了多元性。所谓关系的互惠性是针对两种关系而言的,其含义如下。如果一种关系沿着一个方向"流动",而另外一种关系"反向"流回来,这两种关系就具有互惠性。例如,就"提供情感支持"和"提供物品"这两种关系来说,如果前一种关系的像矩阵是后一种关系的像矩阵的转置矩阵的话,那么在该块模型中就出现了明显的互惠性。另外,"结构均衡性"也可以被表达为两个像矩阵的组合。为了表示结构均衡性,一种正向影响关系的像矩阵会在位置内存在关系,而另外一种负向关系的像矩阵会在各个位置之间存在关系。

二、CONCOR 方法

CONCOR 是一种迭代相关收敛法(convergent correlations 或者 convergence of iterated correlation)。它基于如下事实:如果对一个矩阵中的各个行(或者列)之间的相关系数进行重复计算(当该矩阵包含此前计算的相关系数的时候),最终产生的将是一个仅仅由 1 和 −1 组成的相关系数矩阵。进一步说,我们可以据此把将要计算的一些项目分为两类:相关系数分别为 1 和 −1 的两类。这个现象最初是于 20 世纪 70 年代分别由两个研究小组独立发现的。

具体地说,CONCOR 程序开始于一个矩阵,首先计算矩阵的各个行(或者各个列)之间的相关系数,得到一个相关系数矩阵(C_1)。如前所述,这些相关系数只是表征结构对等性的一种可能的测度。CONCOR 算法的特点是,它把系数矩阵 C_1 作为输入矩阵,继续计算此矩阵的各个行或者各个列之间的相关系数。也就是说,要计算第一个系数矩阵 C_1 的各个行(或者各个列)之间的相关系数。得到的各个"相关系数的相关系数"将构成又一个新的系数矩阵 C_2,然后继续依次计算。最后得到"相关系数的相关系数的相关系数的……矩阵"。

这种迭代过程似乎可以无限计算下去。实际上,经过多次迭代之后,最后的

矩阵中的相关系数值不是 1 就是－1。

例如，假设有一个规模为 14 的网络，计算得到的初始相关系数矩阵为：

表 8.2　初始相关系数矩阵

1.00	-0.20	0.08	0.08	-0.19	-0.19	0.77	0.77	0.77	0.77	-0.26	-0.26	-0.26	-0.26
-0.20	1.00	-0.19	-0.19	0.08	0.08	-0.26	-0.26	-0.26	-0.26	0.77	0.77	0.77	0.77
0.08	-0.19	1.00	1.00	-1.00	-1.00	0.36	0.36	0.36	0.36	-0.45	-0.45	-0.45	-0.45
0.08	-0.19	1.00	1.00	-1.00	-1.00	0.36	0.36	0.36	0.36	-0.45	-0.45	-0.45	-0.45
-0.19	0.08	-1.00	-1.00	1.00	1.00	-0.45	-0.45	-0.45	-0.45	0.36	0.36	0.36	0.36
-0.19	0.08	-1.00	-1.00	1.00	1.00	-0.45	-0.45	-0.45	-0.45	0.36	0.36	0.36	0.36
0.77	-0.26	0.36	0.36	-0.45	-0.45	1.00	1.00	1.00	1.00	-0.20	-0.20	-0.20	-0.20
0.77	-0.26	0.36	0.36	-0.45	-0.45	1.00	1.00	1.00	1.00	-0.20	-0.20	-0.20	-0.20
0.77	-0.26	0.36	0.36	-0.45	-0.45	1.00	1.00	1.00	1.00	-0.20	-0.20	-0.20	-0.20
0.77	-0.26	0.36	0.36	-0.45	-0.45	1.00	1.00	1.00	1.00	-0.20	-0.20	-0.20	-0.20
-0.26	0.77	-0.45	-0.45	0.36	0.36	-0.20	-0.20	-0.20	-0.20	1.00	1.00	1.00	1.00
-0.26	0.77	-0.45	-0.45	0.36	0.36	-0.20	-0.20	-0.20	-0.20	1.00	1.00	1.00	1.00
-0.26	0.77	-0.45	-0.45	0.36	0.36	-0.20	-0.20	-0.20	-0.20	1.00	1.00	1.00	1.00
-0.26	0.77	-0.45	-0.45	0.36	0.36	-0.20	-0.20	-0.20	-0.20	1.00	1.00	1.00	1.00

以该相关系数矩阵为基础，计算该矩阵的各行和各列的相关系数。即进行第一次迭代，得到的结果将是：

表 8.3　初始相关系数矩阵第一次迭代结果

1.00	-.77	0.55	0.55	-.57	-.57	0.95	0.95	0.95	0.95	-.75	-.75	-.75	-.75
-.77	1.00	-.57	-.57	0.55	0.55	-.75	-.75	-.75	-.75	0.95	0.95	0.95	0.95
0.55	-.57	1.00	1.00	-1.0	-1.0	0.73	0.73	0.73	0.73	-.75	-.75	-.75	-.75
0.55	-.57	1.00	1.00	-1.0	-1.0	0.73	0.73	0.73	0.73	-.75	-.75	-.75	-.75
-.57	0.55	-1.0	-1.0	1.00	1.00	-.75	-.75	-.75	-.75	0.73	0.73	0.73	0.73
-.57	0.55	-1.0	-1.0	1.00	1.00	-.75	-.75	-.75	-.75	0.73	0.73	0.73	0.73
0.95	-.75	0.73	0.73	-.75	-.75	1.00	1.00	1.00	1.00	-.77	-.77	-.77	-.77
0.95	-.75	0.73	0.73	-.75	-.75	1.00	1.00	1.00	1.00	-.77	-.77	-.77	-.77
0.95	-.75	0.73	0.73	-.75	-.75	1.00	1.00	1.00	1.00	-.77	-.77	-.77	-.77
0.95	-.75	0.73	0.73	-.75	-.75	1.00	1.00	1.00	1.00	-.77	-.77	-.77	-.77
-.75	0.95	-.75	-.75	0.73	0.73	-.77	-.77	-.77	-.77	1.00	1.00	1.00	1.00
-.75	0.95	-.75	-.75	0.73	0.73	-.77	-.77	-.77	-.77	1.00	1.00	1.00	1.00
-.75	0.95	-.75	-.75	0.73	0.73	-.77	-.77	-.77	-.77	1.00	1.00	1.00	1.00
-.75	0.95	-.75	-.75	0.73	0.73	-.77	-.77	-.77	-.77	1.00	1.00	1.00	1.00

同理，再计算该矩阵的相关系数，即进行第二次迭代，结果为：

表 8.4　初始相关系数矩阵第二次迭代结果

1.00	-.99	0.94	0.94	-.94	-.94	0.99	0.99	0.99	0.99	-.99	-.99	-.99	-.99
-.99	1.00	-.94	-.94	0.94	0.94	-.99	-.99	-.99	-.99	0.99	0.99	0.99	0.99
0.94	-.94	1.00	1.00	-1.0	-1.0	0.97	0.97	0.97	0.97	-.97	-.97	-.97	-.97
0.94	-.94	1.00	1.00	-1.0	-1.0	0.97	0.97	0.97	0.97	-.97	-.97	-.97	-.97
-.94	0.94	-1.0	-1.0	1.00	1.00	-.97	-.97	-.97	-.97	0.97	0.97	0.97	0.97
-.94	0.94	-1.0	-1.0	1.00	1.00	-.97	-.97	-.97	-.97	0.97	0.97	0.97	0.97
0.99	-.99	0.97	0.97	-.97	-.97	1.00	1.00	1.00	1.00	-.99	-.99	-.99	-.99
0.99	-.99	0.97	0.97	-.97	-.97	1.00	1.00	1.00	1.00	-.99	-.99	-.99	-.99
0.99	-.99	0.97	0.97	-.97	-.97	1.00	1.00	1.00	1.00	-.99	-.99	-.99	-.99
0.99	-.99	0.97	0.97	-.97	-.97	1.00	1.00	1.00	1.00	-.99	-.99	-.99	-.99
-.99	0.99	-.97	-.97	0.97	0.97	-.99	-.99	-.99	-.99	1.00	1.00	1.00	1.00
-.99	0.99	-.97	-.97	0.97	0.97	-.99	-.99	-.99	-.99	1.00	1.00	1.00	1.00
-.99	0.99	-.97	-.97	0.97	0.97	-.99	-.99	-.99	-.99	1.00	1.00	1.00	1.00
-.99	0.99	-.97	-.97	0.97	0.97	-.99	-.99	-.99	-.99	1.00	1.00	1.00	1.00

第三次迭代的结果：

表 8.5　初始相关系数矩阵第三次迭代结果

1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
1.00	-1.0	1.00	1.00	-1.0	-1.0	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00
-1.0	1.00	-1.0	-1.0	1.00	1.00	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00

再进行最后一步，把上述只有 1 和 −1 值的矩阵进行重排，将得到如下矩阵：

表 8.6　初始相关系数矩阵第三次迭代结果的重新排列

1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
1.00	1.00	1.00	1.00	1.00	1.00	1.00	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00

我们可以对该矩阵的各个行和列同时进行简化处理，从而得到如下形式：

+1	−1
−1	+1

这样就达到了对所对应的各个行动者进行分区（此表分为两个区，即两个位置），从而简化数据的目的。除此方法之外，我们也可以利用层次聚类对各个行动者进行分类，也可以利用多维量表（MDS）方法对分区进行图像化的表征。利用多种方法对同一个矩阵进行分析可以达到相互补充的目的。

一个网络可以有多个分区，并且每个区本身还可以细分。这也可以利用 CONCOR 对每个子矩阵进行同样的分区操作而得到多个分区。分区越细致，各个区中的行动者就越少。

这就存在一个问题，一个网络应该进行多少次分区？对此没有定论。但是，一般认为，如果经过分区之后，一个区中只有 3 个或者更少的行动者，分区的效果就不好。因此，在最后的结果中，每个区中的行动者最好大于 3 个。

经过多次迭代计算之后，CONCOR 利用树形图（tree-diagram 或者 dendrogram）表达各个位置之间的结构对等性程度，并且标记出各个位置拥有的网络成员。

CONCOR 也可以直接分析多元关系数据以及多值关系矩阵。只是我们要

记住,CONCOR 的分析对象是相关系数矩阵 C_1,它包含的是皮尔逊积距系数,这种系数用来测量各对行动者之间的相似性。至于 CONCOR 在"分区"的时候存在的问题请参见(Wasserman & Faust, 1994:377—380)。

需要指出,上述程序并不能直接给出支持网络中有多少个"块",它需要几个步骤。首先,根据皮尔逊相关系数找出矩阵中有多少个"位置",然后根据密度表以及一定的标准(共有六种,我们将选择其中的一种),确定各个"位置"是 0-块,还是 1-块(参见 Wasserman & Faust, 1994:394—424),从而给出具有高度概括性的"像矩阵",最后还可以画出简化图。

第二节 块模型的应用举例

如果用一个整体网来表征人们之间的相互支持行为,那么如何分析其整体的结构?首先,我们可以计算该整体网络在"帮工"上存在多少个子群(位置)①,然后给出各个位置之间的密度表、像矩阵,最后给出其简化图,从中可以清楚地看出帮工关系网络分为多少个帮派,如表 8.7 所示。

表 8.7 96 户家庭帮工关系网络的密度矩阵表

CONCOR
Density Matrix

	1	2	3	4
	-----	-----	-----	-----
1	0.370	0.092	0.020	0.037
2	0.099	0.388	0.029	0.073
3	0.017	0.025	0.301	0.034
4	0.031	0.103	0.112	0.288

R-squared = 0.161。

这一次的分区情况如表 8.8 所示:

表 8.8 96 户家庭帮工关系网络的分区表

第一个子群的成员有:
1 2 3 4 5 6 7 8 9 34 35 30 37 38 81 28 65 78 85
26 23 50 29 47 41 71 89

① 分析步骤简介如下:在 UCINET 中,沿着 Network→Roles & Positions→Structural→CONCOR 这条路径展开,然后在 CONCOR 中对"帮工关系表"进行分析即可得到下述结果。

第二个子群的成员有：
39 22 12 19 27 64 48 31 21 10 66 49 51 46 45 91 20 82 44
13 14 79 11 67 77
第三个子群的成员有：
84 72 15 42 57 88 32 60 76 56 63 70 86 18 73 68 69 58 59
96 61 74 75 94 17 33 55 83 87 43 95 93
第四个子群的成员有：
16 80 54 52 62 36 25 92 90 40 53 24

整个网络的密度值经过计算为 0.1277，得到的像矩阵如表 8.9 所示。

表 8.9 96 户家庭帮工关系的像矩阵表

	1	2	3	4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1

分析得到的完全是一个对角线矩阵，由此得到如下结论：村民们的帮工网络分为四个子群（subgroup），并且完全是自反性的（self-reflexive）。也就是说，总体上，帮工现象出现在子群的内部成员之间，而各个子群之间基本上不相互帮工。可见，在这种关系网络中，帮派性表现得很明显。

总之，尽管涉及帮工网络的家庭成员有 96 个，但是从整体上说，其简化图却非常简单，如图 8.4 所示。

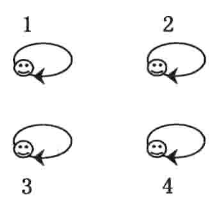


图 8.4 96 户村民的帮工关系简化图

图中的每一个点（头像）代表一个子群，每个点上面带箭头的小圆圈，表示的是关系从该点“发出”，又回到本点。在各个点之间没有连线，表示在各个子群之间不存在相互帮工的情况。也就是说，这四个子群之间不存在帮工方面的来往，但是这不否认子群之间在其他方面（如借钱、贷款、借小物品等）有交往。

本 章 小 结

块模型是一种用来揭示整体网结构的方法,本章介绍了块模型的含义、构建以及解释方法,介绍了两个位置和多个位置中的像矩阵。分析块模型的方法是CONCOR法,本章结合实际例子展示了块模型的应用。

第九章

个体网研究:结构洞与中间人

UCINET 把所有的数据都看成是矩阵,其中的 1-模数据往往是整体网数据。我们可以从中抽取出每个点的个体网数据,进而对个体网的多个特征进行计算和分析,其中结构洞(structural holes)和中间人角色(brokerage roles)是比较重要的两种特征,下面加以探讨。

第一节　局部桥和结构洞

一、局部桥的含义及测量

格兰诺维特在其名篇(Granovetter, 1973)中提出了"弱关系的强度"假设。他认为,弱关系的重要性在于,它在群体、组织之间建立了纽带关系,传递着信息,而强关系维系着组织内部的关系;弱关系在寻找工作信息的时候也具有重要意义。除此之外,弱关系的重要性在于,它可以使得一个更大的网络达到结构上的凝聚性。人们保持的弱关系常常比强关系多。格氏指出,强关系常常处于群体内部。因此,把一个网络中的不同群体结合在一起的关系主要是弱关系,不管该关系是局部桥还是桥。布劳也指出了这个观点的重要性:"亲密的关系容易局限在亲近的社会圈子里,因此,它们会使社会发生分崩离析。各个群体在社会中的整合并不取决于人们之间的强有力的联系,而只取决于他们之间松散的联系,因为松散的社会联系会大大地超越亲密的社会圈子,从而使各群体建立起社会关系。"(布劳,1991:126—127)"过量的内群体联系会危害社会整合,因为它易于导致社会的分裂。"(布劳,1991:128)

(一) 桥、局部桥的含义

什么是桥(bridge)? 什么是局部桥(local bridge)? 在一个关系网络中,"桥"指的是这样的两点之间的关系:如果去掉该关系,整个网络将分为两个独立的子网络。或者说,如果两个群体之间仅仅通过唯一关系连在一起,则称此关系

为“桥”(Granovetter，1973:1364)。而“局部桥”的定义稍有不同:“如果 n 大于 2 并且是连接两点的最短途径的长度(除了已存的关系之外)，该现存关系就是度数为 n 的局部桥。”(Granovetter，1973:1365)可见，局部桥的度数至少为 3。

在一个现实网络中，桥和局部桥往往是信息的通道，资源交换的关节点，控制资源的重要任务，因而具有重要意义。在图 9.1 中存在多个桥，如 AB，AC，AD，BE，BF，CG，DH，DI 都是桥。

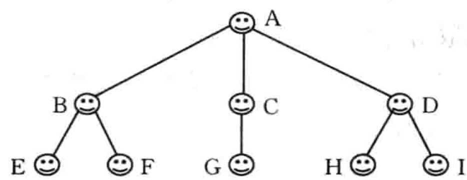


图 9.1　9 点图中的桥

在图 9.2 中不存在桥，却包含多个局部桥。但是，P_1-P_2不是局部桥，因为除了二点之间现存的直接关系之外，其余“路径”的最短长度为 2，因为它不满足“局部桥”的定义。P_1-P_7则是度数为 7 的局部桥，因为，根据“局部桥”定义，如果不考虑 P_1 和 P_7 之间的直接关联而分析从 P_1 出发需要经过多少“途径”(paths)才可以达到 P_7，会发现至少要经过 7 个途径，所以，P_1-P_7 是局部桥，其度数为 7。

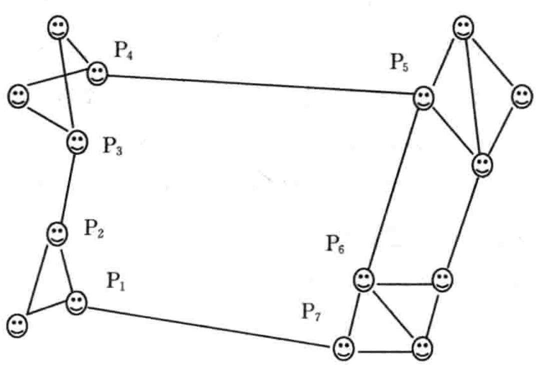


图 9.2　桥和局部桥

(二) 局部桥的测量

如何计算一个网络中包含的桥和局部桥？这取决于关系的内容和对关系

“强度”的界定。格兰诺维特认为：“一个关系的强度有可能是时间的多少、情感的紧密性、亲密性以及互惠性的服务这些刻画关系特点的属性构成的线性组合。”(Granovetter，1973：1361—1364)在实际分析中，他又把不对称关系看成是“弱关系”，把对称关系看成是“强关系”。马岑(Mardsen，1990)进一步认为，对关系强度的测量有两个方面：关系维持的时间和关系的深度。而关系的“远近”是测量深度的最好指标，用频次和接触的时间作为强度的指标则有问题。

格兰诺维特指出，“所有的局部桥都是弱关系”(Granovetter，1973：1364)。在现实中，该命题为真的可能性不大，因为毕竟有的局部桥可能不是弱关系。但是，我们有理由提出如下假设：大多数局部桥应该是弱关系。下面用一个村落(可称之为法村)的村民之间的社会支持数据展示桥数量的计算(本部分主要取自刘军，2006a：181—187)并对该假设进行检验。我们用“关系远近”而不用关系的对称性或者“近亲”关系等来测量村民之间关系的强度。以家庭为调查单位，分析各个家庭之间所有可能存在的如下“紧密性”的远近关系：(1)没有来往；(2)关系一般；(3)关系较好；(4)关系很好。

检验方法：分别结合“亲友关系”、“远近关系”，利用 UCINET 中的矩阵代数(Matrix Algebra)算法算出局部桥的个数，计算出在这些桥中有多少为弱关系。

首先分析在“亲友关系”中存在的桥有多少。在此数据中，我们定义亲属关系为强关系，朋友关系为弱关系，计算结果如下：桥的总数有 194/2 = 97 个，这些关系都是弱关系(具体的操作化过程，参见刘军，2006：186)。也就是说，弱关系占总数的 100%，强关系一个也没有。可见，在亲属关系中，“桥”都是“弱关系”，我们的假设是完全成立的。这个结论也恰恰验证了格兰诺维特的上述论断。对上述结果的解释如下：在法村，亲友关系可以分为四个派系。各个派系之间不是绝对隔离的，而是通过一定的中介关系连在一起。这些中介关系(桥)一共有 97 个，它们都是弱关系。在村民家庭之间的“远近关系”问题上也类似，经过计算(各种指令的具体计算过程参见刘军，2006：186)，得到如下结果：

桥总数目为 88/2＝44 个桥。其中，强关系有：10/2＝5 个；弱关系有 78/2＝39 个。我们把针对两种关系的桥的计算结果汇总，如表 9.1 所示。

表 9.1 亲友关系和远近关系中的局部桥所占比例

关系的类型(矩阵规模)	桥的总数	强关系	弱关系	弱关系所占比例
亲友关系(103×103)	97	0	97	100%
远近关系(103×103)	44	5	39	89%

从结果中可以看到,"亲友关系"和"远近关系"中的弱关系所占比例都很高,分别是 100%和 89%。实际上,这是因为这两种关系之间具有高度的相关性。但是,这种相关性是否在统计意义上显著?利用 UCINET 中的 QAP 命令进行检验,结果显示,在统计的意义上,二者之间的关系是显著的。或者说,两个村民越是亲友关系,他们之间相处地"越近"。因此,在"桥"上,"远近"关系也表现出"亲友"关系类似的性质,大多数桥是弱关系。因此,研究假设"桥多数是弱关系"得到了统计描述意义上的检验。

二、结构洞

(一)结构洞的含义

什么是结构洞(structural holes)?伯特用结构洞来表示非冗余的联系,认为"非冗余的联系人被结构洞所连接,一个结构洞是两个行动者之间的非冗余的联系"(Burt, 1992:18)。例如,在图 9.3 中,"自我"与 A, B, C 中的任意两者之间的关系结构就是一个结构洞。因为,A 和 B 都与"自我"有关系,但是二者之间却不存在关系,相当于有一个空洞(hole)。"自我"如果希望把信息传递给 A 和 B,需要分别通知;而在图 9.4 中,"自我′"仅把信息传递给 A′即可,因为 A′可以把信息传递给 B′。也就是说,对于"自我"来讲,A 与"自我"的关系和 B 与"自我"的关系是非冗余的;而对于"自我′"来讲,A′与"自我′"的关系和 B′与"自我′"的关系则是冗余的。"自我"便是结构洞的中间人或者占据者。伯特认为,结构洞能够为其占据者获取"信息利益"和"控制利益"提供机会,从而比网络中其他

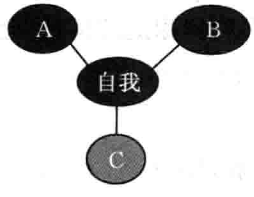


图 9.3 有结构洞的信息流动网

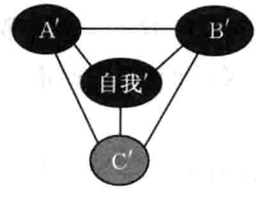


图 9.4 关系闭合的信息流动网

位置上的成员更具有竞争优势。

仅仅从关系缺失的角度并不能完全说明结构洞。判断结构洞的标准有两个：凝聚性(cohesion)和对等性(equivalence)。凝聚性的含义是，如果一个行动者的两个联络人之间存在直接的关系，凝聚力加大，冗余性也增强。例如，在派系中就不存在结构洞。对等性的含义在于考虑到了“自我”与其网络成员之间的间接关系。如果两个行动者与网络中的同一群人之间共享同样的关系，我们便说这两个人(如图 9.5 中的 A 和 B)之间是结构对等的。他们之间可能没有直接的联系，但是每个人的关系网却可能一样，因此从“自我”的角度看，这两个人(如A、B)提供的信息是冗余的。自 1992 年以来，伯特就致力于结构洞理论的推广研究，在此方面发表了大量著述。

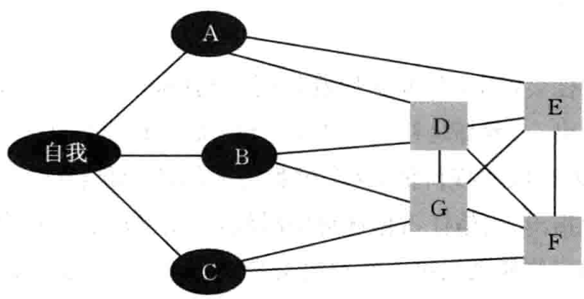


图 9.5　结构对等性

可见，格兰诺维特所说的“弱关系”指的是两个行动者之间关系的“强弱”这种具体性质，不管如何界定关系的强弱。“桥”或者“局部桥”都是一种特殊的关系，尽管大多数情况下它们都是弱关系。而伯特所说的结构洞则代表由至少三个行动者之间关系构成的一种特殊结构，这种结构可能为中间人带来利益。因此，这种结构就会成为社会资本，其中间者扮演着中介人(broker)的角色。

结构洞这种结构能够为中间人提供什么好处？这要看网络的性质，也要看网络中的行动者是什么。如果行动者是个人，那么这种结构很可能为该人提供非冗余的信息，从而为自己带来职位的晋升、工资的增长或者声望的提高等(Burt，2004)；如果行动者是社区，这种结构可能为社区的发展提供机遇；如果是高校，这种结构可以为学校声望的提高作出贡献。无论如何，结构洞会成为社会资本，对占据者来说有重大意义。

(二) 结构洞成为社会资本的机制何在?

为什么结构洞的这种经纪行为(brokerage)会成为社会资本?其机制是什么?关键在于中间人在两个不同的群体之间建立关系。如伯特所言,这是因为“在群体内的观念和行为要比在群体之间的观念和行为更具有同质性,所以跨群体之人会更熟悉另类想法和行为,因而会给他们更多的观念选择的机会……这种视野优势恰恰是经纪行为变成社会资本的机制。”(Burt, 2004:349)

科学家的学术创新也类似,他们往往是通过圈外之人获得灵感的,这是因为通过圈外之人可以建立结构洞,结构洞对于获得新观念具有重要意义。学术研究也类似,通过博览群书(特别是阅读其他学科的著述),接触各种相关学科的知识,与各个相关学科领域的学者深入交流,“新思想”自然会不断产生。这在一定意义上可以解释为什么学术上的“近亲繁殖”很难带来学术创新,却带来学术小派系的繁荣。

结构洞成为社会资本,并不意味着封闭结构(network closure)不能成为社会资本。伯特指出,二者都是社会资本,这不矛盾。这是因为,通过把结构洞桥接在一起,中间人可以实现价值,这个过程需要信任,而信任恰恰是靠网络的封闭性保证的。他指出,当信任占据优势的时候,封闭性就成为社会资本。结构洞的社会资本依赖于信任来提供由经纪机会带来的价值。因此,在现实生活中,行为的表现是“群体之外的经纪行动和群体之内的封闭关系的产物”(Oliver, Yuval Kalish & Gad Yair, 2007:334)。

(三) 结构洞理论的贡献

结构洞理论的贡献在于它把散布在各种文献中的相关思想(如弱管理的强度、渔利者、局部桥思想、网络交换论等)整合成为一种一致的理论框架,指出了这种结构中的中介者具有信息优势和控制优势,因而有广泛的应用。除此之外,笔者认为该理论还有两个相互关联的贡献:该理论所倡导的网络行动观沟通了微观和宏观的连接(Burt, 1992: 第5章;Burt, 1982);强调了三方关系结构的重要性,而三方关系揭示了整个社会世界的核心架构。

1. 结构洞理论倡导网络行动观,沟通了宏观与微观之间的桥梁。

总的来说,宏观行动观起源于孔德所创立的实证主义传统,这种行动观有一些预设,如客观性假设、无“污染”假设、社会规律性假设和利用自然科学方法研究社会问题等。而微观行动观的代表人物很多。例如,加芬克尔发现了日常生活的实践活动的权宜性、局限性和索引性,这表明了社会世界具有不确定性,行

动也不是由规则预先决定的,它是一个在局部场景中由行动者通过权宜性的努力后构成、创造的结果。而在现象学传统中,舒茨认为,社会学的任务是研究行动者是如何赋予行动以意义的,其研究主题已经不是“行动”而是“意义”。相比之下,“网络行动观”可以沟通宏观视角和微观视角之间的“鸿沟”,因为它既看到了行动中个人的努力,也看到了规范结构的力量。在网络数据中没有什么微观和宏观的分析,尽管数据可能是在微观系统的基础上收集到的。当然,这里的行动者可以是个人,也可以是非正式群体或者正式的企业等。所以,行动者即可以是微观水平的,也可以是宏观水平的。例如,科层体制下的雇员、社区精英、无形学院中的精英、社区中的大型企业甚至是处于世界体系中的国家。

2. 结构洞理论强调了三方关系结构的重要性。

科尔曼认为,社会关系可以分为两类:简单关系(即二人关系)和复杂关系(即三人关系)。复杂关系的维持必须依赖于第三者。这种关系常常是正式组织得以生存的基础。复杂关系绝不是二方关系的简单叠加,它与二人关系有本质的区别,结构洞至少涉及三个行动者之间的关系。在《道德经》中有如下表述:“道生一,一生二,二生三,三生万物。”只有在“三”的基础上,万物才得以生成。结构洞理论的贡献在于它超出两个行动者之间关系的“强弱”层面,关注到了至少三个行动者之间关系的结构层面,而“三方关系”深刻地揭示了整个世界的架构,恰恰是整个社会得以建构的基石,是社会团结的基础(刘军,2006b)。

(四)结构洞的测量

结构洞的计算比较复杂,总的来说存在两类计算指标。第一类是伯特(Burt, 1992)本人给出的结构洞指数,第二类是中间中心度指数。后一类还包括弗里曼(Freeman,1979)给出的针对整体网的中间中心度指标及其推广形式(Brandes, 2008),以及艾弗雷特等学者(Everett & Borgatti, 2005)给出的针对个体网数据的个体中间度指标,下面分别加以探讨。

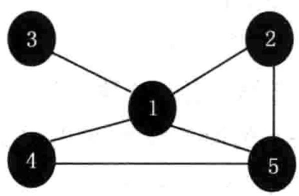
1. 伯特的结构洞指数。

伯特的结构洞指数要考虑四个指标:(1)有效规模(effective size);(2)效率(efficiency);(3)限制度(constraint);(4)等级度(hierarchy),其中第三个指标最重要。

(1) 一个行动者的有效规模等于该行动者的个体网规模减去网络的冗余度(redundancy),即有效规模等于网络中的非冗余因素。行动者 i 的有效规模可用公式表示为:

$$\sum_{j}(1-\sum_{q}p_{iq}m_{jq}),\ q\neq i,\ j$$

其中,j 代表与自我点 i 相连的所有点,q 是除了 i 或 j 之外的每个第三者。括号内部的量 $p_{iq}m_{jq}$ 代表在自我点和特定点 j 之间的冗余度。其中 p_{iq} 代表行动者 i 投入到 q 的关系所占比例,m_{jq} 是 j 到 q 的关系的边际强度(marginal strength),它等于 j 到 q 的关系取值除以 j 到其他点关系中的最大值。当然,对于二值网络来说,最强值就是 1,因此 m_{jq} 总是 1 或者 0。乘积 $p_{iq}m_{jq}$ 的总和测量的是 i 与 j 的关系相对于 i 与其他关系人的关系的比例。图 9.6 代表一个 5 点图及其邻接矩阵,下面将结合该图进行分析。



	1	2	3	4	5
1	0	1	1	1	1
2	1	0	0	0	1
3	1	0	0	0	0
4	1	0	0	0	1
5	1	1	0	1	0

图 9.6 一个 5 点图及其邻接矩阵

实际上,鉴于上述计算比较麻烦,薄伽提(Borgatti, 1997)给出了简单的算法,即 i 的冗余度就等于该点所在的个体网成员在网络中的平均度数(连接到中心点的线不计算在内)。例如,对于点 1 来说,与之相连的点有 2, 3, 4, 5。它们的度数分别是 1, 0, 1, 2。因此,点 1 的冗余度如表 9.2 所示。

表 9.2 点 1 的网络成员及其密度

点 1 的个体网成员	2	3	4	5	均 值
度 数	1	0	1	2	4/4=1

所以,点 1 的冗余度是 1,有效规模是 4－1=3。要想计算点 5 的有效规模,需要计算出该点的冗余度 1.33(如表 9.3 所示),由于点 5 的个体网络的规模是 3,因而点 5 的有效规模是 3－1.33＝1.67。

表 9.3 点 5 的网络成员及其密度

点 5 的个体网成员	1	2	4	均 值
度 数	2	1	1	4/3 = 1.33

可见,一个点的冗余度就是该点所在个体网的平均度数。因此,我们可以简化计算行动者 i 的冗余度为 $2t/n$,其中 t 是 i 的个体网络中的关系数(不包括与中心点相连的关系数),n 是 i 的个体网络规模(不包括自我点)。因此,i 的有效规模是 $n-2t/n$ (Borgatti, 1997:37)。

(2) 一个点的效率等于该点的有效规模与实际规模之比。例如,点 1 的效率为 3/4。

(3) 限制度(constraint):一个人受到的"限制度"指的是此人在自己的网络中拥有运用结构洞的能力。按照伯特的说法,"你自己的机会受到的限制取决于你曾经投入了大量网络时间和精力的另外一个接触者 q 又在多大程度上向接触者 j 的关系投入大量的精力"(Burt, 1992:54)。由此引出限制性的操作化定义:行动者 i 受到 j 的限制度指标为 $C_{ij}=(p_{ij}+\sum_q p_{iq}p_{qj})^2$,可表述为 C_{ij}=直接投入(P_{ij})+间接投入($\sum_q p_{iq}p_{qj}$)。其中 p_{iq} 是在行动者 i 的全部关系中,投入到 q 的关系占总关系的比例,其含义如图 9.7 所示。

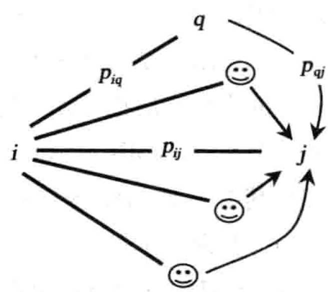


图 9.7　限制度计算例示

这个表述稍显复杂。对于二值数据来说,可以更清楚地表述如下。如果行动者 i 的个体网规模为 n_i,则 $p_{ij}=p_{iq}=\frac{1}{n_i}$。关键看 p_{qj}。根据定义,p_{qj} 是在行动者 q 的全部关系中,投入到 j 的关系占总关系数的比例。在二值网络中,如果行动者 q 的个体网规模为 n_q,则 $p_{qj}=\frac{1}{n_q}$。这样看来,公式 $C_{ij}=(p_{ij}=\sum_q p_{iq}p_{qj})^2$ 可以简化为 $C_{ij}=\left(\frac{1}{n_i}+\sum_q \frac{1}{n_i}\frac{1}{n_q}\right)^2=\left(\frac{1}{n_i}\right)^2\left(1+\sum_q \frac{1}{n_q}\right)^2$。这个公式清楚地展示了限制度计算的实质。

间接关系总量 $\sum_q p_{iq}p_{qj}$ 反映了罗纳德·伯特(Ronald Burt)建构指数的逻辑。首先,由该公式可见,我们关注的只是 i 和 j 通过一个中间人 q 建立的传递关系及其控制力。如果读者同时还关注 i 和 j 通过两个甚至多个中间人建立的传递关系或控制关系,那么上述公式就变得很复杂,也没有太大的现实意义。其次,$p_{iq}p_{qj}$ 反映的是 p_{iq} 和 p_{qj} 之间的"相乘"关系,其实这里的潜在前提性假设是:

关系的投入强度随着关系链的增长而呈非线性的降低趋势,因而用这二者之积来表达。

限制度指数 C_{ij} 的最小值为 p_{ij} 的平方(即 j 与其他点都不相连),最大值是 1(如果 j 是 i 的唯一联络人)。在这个公式中,取遍所有的联络人 j 得到的总和就测量了行动者 i 在网络中的创业机会受到的总限制性(aggregate constraint)(Burt, 1992:55)。

以图 9.6 中的"1"为例。与之接触的"5"对 1 的控制最强,因为在"1"的个体网成员当中,5 与网络中其他者之间的联系最多。5 对 1 的要求因而将是 1 最难以回避的。3 对 1 的限制性最小,因为它与 1 的网络中的其他成员之间完全隔离。因此,3 的要求具有最大的可协商性。

上述限制度是针对单个行动者来说的。如果在开始收集数据的时候就根据一定的先验标准把行动者进行了分组,那么我们就可以针对每一组计算出一个独断值(oligopolistic value)O_j,对该值的测量至少有三种(Burt, 1992:63—65),其中一种与组内结构洞的数量成反比(Degenne & Forse, 1999:127—128)。如果一个子群体是派系,则该派系的独断值为 1,如果一个子群体的规模是 n_j,并且包含的点都是孤立点,则该派系的独断值达到最小值 $1/n_j$。由此引申出第二类结构洞。

在这种结构洞中,j 所在群体的结构洞数目越多,j 对其网络外的 i 的限制力越小。上文界定了 c_{ij},这个值乘以 O_j 便可得到一般意义上的 j 对 i 的限制值,即 $C_{Gij}=c_{ij}O_j$。群体之间的结构洞越少,c_{ij} 越大;群体内部的结构洞越少,O_j 越大。因此,二者之积 C_{Gij} 也随着群体内部结构洞和群体之间结构洞的减少而增加。在图 9.8 中,$C_{Gij}<C_{Giq}$ (Degenne & Forse, 1999:128)。有关这方面的例子请参见(Burt, 1992:63—65,特别是图 2.5)。

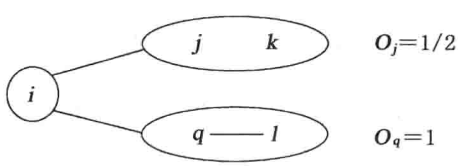


图 9.8 第二类结构洞例示

(4) 等级度(hierarchy)。在伯特看来,等级度指的是限制性在多大程度上集中在一个行动者身上。其计算公式为(该公式被称为科尔曼—泰尔失序指数,

Coleman-Theil disorder index)：

$$H = \frac{\sum_{j} \left(\frac{C_{ij}}{C/N}\right) \ln\left(\frac{C_{ij}}{C/N}\right)}{N \ln(N)}$$

其中 N 是点 i 的个体网规模。C/N 是各个点的限制度的均值，公式的分母代表最大可能的总和值。当一个行动者的每个联络人的限制度都一样的时候，该测度达到最小值 0。反之，当所有的限制都集中于一个行动者的时候，该值就达到最大值 1。或者说，一个点的等级度越大，说明该点越受到限制，反之亦反。

需要指出的是，伯特所给出的结构洞指数测量的是个体在网络中的受限制度，他所依据的网络是个体网而不是整体网。类似的思路也可以在整体网络中实现，也就是说，我们也可以按照上述思路计算整体网络中的限制度指数等结构洞指标，这时候需要考虑到距离"核心点"长度超过 2 的途径，后文将有针对整体网数据的结构洞指标的计算。除此之外还有一种思考问题的方式，也就是说，如果一个行动者在网络中居于很多人中间，即处于许多路径上，我们就可以说此人拥有较多的结构洞。因此，我们可以计算一类"中间度"指标（参见本书第五章），用它来测量结构洞。这个指数的计算可分别结合整体网和个体网来探讨。

2. 伯特的结构洞指数测量例示。

下面，我们以图 9.6 的数据为例展示伯特结构洞指数的计算过程。在社会网络分析软件 UCINET 中（Borgatti, Everett & Freeman, 2002），沿着 Network→Ego-networks→Structural Holes 这条路径，在 Input dataset（输入数据）一项中选出在 UCINET 中事先构建好的图 9.6 数据矩阵。在 Method（方法）一项中有两个选项。如果数据是整体网，则选择"Whole Network model"；如果是个体网，则选"Ego Network model"。这里选择"Whole Network model"，因为我们关注是在整体网中每个点的结构洞情况（当然，如果把图 9.6 看成是一个以"1"为核心的个体网，则选"Ego Network"，会出现表 9.4 右侧所示的计算结果，可以看出这两个表是不同的）。点击 OK 即可计算出该网络中每个点的结构洞指标，如表 9.4 所示。

表 9.4 把图 9.6 分别看成是整体网(左)和个体网(右)进行结构洞分析的结果

左(整体网)

STRUCTURAL HOLES

Input dataset: H:\5 nodes holes

Method: Whole Network

Dyadic Constraint

	1	2	3	4	5
1	0.00	0.11	0.06	0.11	0.25
2	0.44	0.00	0.00	0.00	0.39
3	1.00	0.00	0.00	0.00	0.00
4	0.44	0.00	0.00	0.00	0.39
5	0.44	0.17	0.00	0.17	0.00

Structural Hole Measures

	EffSize	Efficie	Constra	Hierarc
1	3.000	0.750	0.535	0.092
2	1.000	0.500	0.835	0.003
3	1.000	1.000	1.000	1.000
4	1.000	0.500	0.835	0.003
5	1.667	0.556	0.792	0.099

右(个体网)

STRUCTURAL HOLES

Input dataset: H:\5 nodes holes

Method: Ego Network — connections 2 links beyond ego are ignored

Dyadic Constraint

	1	2	3	4	5
1	0.00	0.11	0.06	0.11	0.25
2	0.56	0.00	0.00	0.00	0.56
3	1.00	0.00	0.00	0.00	0.00
4	0.56	0.00	0.00	0.00	0.56
5	0.44	0.20	0.00	0.20	0.00

Structural Hole Measures

	EffSize	Efficie	Constra	Hierarc
1	3.000	0.750	0.535	0.092
2	1.000	0.500	1.125	0.000
3	1.000	1.000	1.000	1.000
4	1.000	0.500	1.125	0.000
5	1.667	0.556	0.840	0.074

对左侧表的结果解释如下:先看左上表,该表表示任何两个点之间的限制度。可以看出,点 5 对点 1 的限制度最大,达到 0.25。而点 3 受到点 1 的限制度为 1,这是限制度的最大值,说明点 3 受到点 1 100%的限制。对其他值的解释也类似。再看左下表,该表给出了每个点的有效规模、效率、在整体网络中的总限制度以及等级度。可以看出,点 1 的有效规模是 3,受到的总限制度达到0.535(该值约等于第一个表中第一行值的总和);而点 3 受到点 1 的限制度达到 1,其等级度也达到最大值 1,这说明点 3 处于最高的等级。

右侧表给出的是根据个体网资料计算得到的结果。可以看出左右两表之间的差异。对于点 1 和点 3 来说,两个表的计算结果是一样的。但是对于其他 3 个点来说,计算的结果不同。原因在于无论把图 9.6 看成是整体网还是个体网,点 1 和点 3 的个体网成员都是一样的,计算结果当然也相同。但是对于其他三个点来说,情况有变,因此计算结果稍有差异。研究人员应该结合自己研究的网络的性质选择采用整体网还是个体网进行计算。

实际上，如果研究人员面对的是个体网资料，也不用根据上述步骤一个一个地进行计算，因为这样太麻烦。这时候，可以采用两种办法。一种办法是先制作一个大的数据矩阵，即将多个个体网数据构建成为一个大的数据矩阵，这是一个多元数据矩阵，然后利用上述思路进行计算。第二种办法是利用专门分析个体网资料的程序 E-net（可从 http://www.analytictech.com/网站上免费下载）。该程序由肯塔基大学（University of Kentucky）斯蒂芬·薄加提（Steve Borgatti）教授编写[①]，所分析的数据既可以包括属性资料（如性别、年龄、职业等），也接受关系资料（包括自我与他人的关系以及他人与他人之间的关系）。该程序可以分析个体网数据的很多性质，如个体网的构成（如成员中包含多少女性）、同质性和异质性，同时也能够计算出每个个体的结构洞指标。其具体计算步骤及过程展示这里从略。

3. 中间中心度指标及其推广。

(1) 整体网络中个体的中间中心度。

① 弗里曼给出的中间中心度指标。鉴于结构洞的中间人往往占据重要地位，因此，我们可以用弗里曼（Freeman，1979）给出的中间中心度（betweenness centrality）指数作为结构洞指数，用它来测量行动者对资源的控制程度。在本书第五章即“中心性和影响力”一章中已经对中间中心性进行了介绍，下面进行一些补充。

中间中心度是最常见的一种中心度指数，是社会网络分析的一个核心概念。下面介绍该指数的描述性含义和操作化公式。

在一个整体网中，如果一个点处于许多其他点对的捷径上，我们就说该点具有较高的中间中心度，它可能起到重要的“中介”作用，处于网络的中心，拥有较多的结构洞。中间中心度的描述性定义如下：一个点 Y 处于 X 和 Z 之间的中间度指该点处于此点对的捷径上的能力，具体用“中间性比例”来刻画，其定义为：经过点 Y 并且连接两点 X 和 Z 的捷径与这两点之间的全部捷径数之比，它测量了 Y 在多大程度上位于 X 和 Z 的“中间”。一个整体网中有很多点对，一个点 Y 相对于网络中所有点对的“中间比例”的总和便是该点在整体网中的“中间中心度”。这是对中间中心度的描述性说明，其公式化表征如下。

对于有向图 $G=(V, E)$ 来说，我们界定 V 是点的集合，$E \subseteq V \times V$ 是有向边的集合。对于一个有向边 $e=(u, v) \in E$ 来说，我们分别称 u 和 v 为该边的

① 感谢英国威斯敏斯特大学马丁·艾弗雷特（Martin Everett）教授向笔者分享该信息。

头和尾。从点 s 出发,到达点 t 的途径(path)可能有多个,记做 (s, t)-途径,其包含的线数叫做该途径的长度,其中最短者叫做该途径的距离,记做 $\mathrm{dist}(s, t)$。用 $\sigma(s, t)$表示最短的 (s, t)-途径数目,并且令 $\sigma(s, t|v)$表示经过点 v 的最短 (s, t)-途径的数目。如果 $s=t$,则令 $\sigma(s, t)=1$,并且如果 $v \in (s,t)$,则令 $\sigma(s, t \mid v)=0$。这样便可以用公式表示点 $v \in V$ 的中间中心度 $c_B(v)$(同时定义 $0/0=0$):

$$c_B(v) = \sum_{s,\, t \in V} \frac{\sigma(s,\, t \mid v)}{\sigma(s,\, t)}。$$

这就是弗里曼给出的中间中心度表达式,其具体计算实例参见(Freeman, 1979;刘军,2004:122—125)。标准化的中间中心度的取值介于 0 和 1 之间,其值越大,说明对应的点越是中间人,结构洞越多,越居于整个网络的核心;反之亦反。

② 整体网中间中心度的推广。上述中间中心度指数是针对整体网来测量的,其前提假设很简单,没有考虑到关系的强度、方向以及关系的多元度等。可想而知,在很多复杂情况下,这种计算方法可能不恰当。考虑到这些问题,布兰迪斯(Brandes, 2008)探讨了中间中心度的各种推广形式。

第一种推广的思路如下:在弗里曼的中间中心度指标的界定中,各个点对之间的关系不依赖于自身,却依赖于中间点,有时候这可能不太恰当。因此,第一个推广形式便是将起点和终点计算在内,这时候需要修改一些界定,即当 $v \in (s, t)$ 的时候,将原先界定的 $\sigma(s, t \mid v)=0$ 修改为 $\sigma(s, t \mid v)=1$。这样的话,每个 $v \in V$ 的中间中心度指标都会增加,即增加 v 可达的点数再加上可达到 v 的点数,这样可以得到第一种推广形式。

第二种推广考虑到了如下事实:在一对点之间的关系中,有时候我们有理由相信,只有关系的最初起点(或最终点)能够控制资源,据此对中间中心度进行推广。

第三种推广的思路如下:常规的中间中心度指标考虑两点之间的最短捷径,不关注其长度是多少,由于对于某些关系(如朋友)来说,太长关系链的意义不大,因此,薄伽提等人(Borgatti & Everett, 2006)界定了一个点的 k-中间度(k-betweenness):两点之间长度不超过 k 的捷径计算在内,超过者不计算,用公式表示为:

$$c_{B(k)}(v) = \sum_{s,\, t \in V;\, dist(s,\, t) \leqslant k} \frac{\sigma(s,\, t \mid v)}{\sigma(s,\, t)}$$

第四种推广的思路如下:有时候我们有理由相信,途径越长,其控制的能力和价值越小,由此引申出另外一种推广方法,即将全部最短途径用其长度的倒数进行加权,表示为:

$$c_{B(dist)}(v)=\sum_{s\neq t\in V}\frac{1}{dist(s,\ t)}\cdot\frac{\sigma(s,\ t\mid v)}{\sigma(s,\ t)}$$

还有一些推广形式,如针对边中心度、群体中心度、2-模网络的中心度、多值网络和多元网络的中心度等。所有这些中间中心度指数都可以用计算机实现,其算法请参见布兰迪斯(Brandes, 2008)。

可见,中间中心度指数的这些推广形式的适用范围很广。由于伯特关注的网络多数仅仅是二值无向关系网,并且是个体网,没有考虑到其他条件(如途径的长度、距离、多值网、2-模网等),因此上述中间中心度指数在很大程度上推广了伯特的结构洞指数。

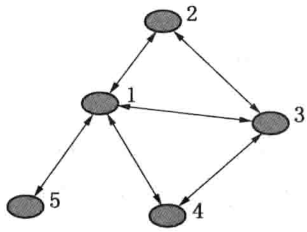
(2) 个体网中"核心点"的中间中心度。

伯特探讨的网络限制性指数等主要适用于测量个体网的结构洞,而上述中间中心度的定义主要适用于整体网。实际上,针对个体网也可以测量核心点的中间中心度,这便是艾弗雷特等学者(Everett & Borgatti, 2005)探讨的问题。

测量中心度的一个重要目的便是找出网络中的重要行动者。在整体网中,可以考虑对行动者的中间中心度进行标准化处理,其目的是为了比较。但是在个体网中,一般不对个体网的一些指标进行标准化,因为这样可能产生一些问题,例如网络的规模各不相同,无法比较。可以认为,个体网规模越大,个体在自身的个体网之外越可能处于其他行动者的中间,结构洞也越多。

个体网中核心点的中间度计算比较简单。首先,根据个体网的定义,我们只需要计算一个点(即核心点)的中间中心度。其次,网络中捷径的长度不超过 2。每一对非邻接的点的捷径长度都是 2,我们只需要考虑这样的非邻接点即可,因为长度为 1 的捷径对中间度没有贡献。如果矩阵 A 是个体网图 G 的邻接矩阵,那么 $A^2_{i,j}$ 便包含连接 i, j 的长度为 2 的线路(walk)数。当 $i\neq j$ 的时候,这个线路必然是途径(path)。我们只需要计算非邻接的点对之间长度为 2 的途径数,因为这些才是捷径。既然 A 是邻接矩阵,$J-A$ 便是非邻接矩阵(其中矩阵 J 是一个其全部元素的取值都是 1 的矩阵)。因此,可以证明,矩阵 $A^2[J-A]_{i,j}$ 给出的恰恰是连接 i 和 j 的长度为 2 的捷径(geodesic)数,那么该矩阵中各个非零元素(对角线除外)倒数之和便是核心点的中间度测度。

例如,假设有一个个体网图及其邻接矩阵如图 9.9 所示,我们现在展示如何计算该网络的核心点“1”的中间度。



$$\text{邻接矩阵 } A_{5\times 5}=\begin{bmatrix}0&1&1&1&1\\1&0&1&0&0\\1&1&0&1&0\\1&0&1&0&0\\1&0&0&0&0\end{bmatrix}$$

图 9.9　一个 5 点个体网图及其邻接矩阵

由于这是一个对称矩阵,我们只需要考虑主对角线上方的元素即可。在 UCINET 软件的矩阵算法(Tools→Matrix Algebra)中,首先计算矩阵 A 与自身之积(production),即需要键入命令 AA = prod(A, A),回车。其次计算矩阵 AA 与矩阵 J-A 的对应元素之积(multiplication),令其等于 C,即键入命令 C = mult(AA, (J—A)),回车后用 display 命令即可展示出计算结果如下所示:

$$A^2[J-A]=\begin{bmatrix}4&0&0&0&0\\0&2&0&2&1\\0&0&3&0&1\\0&2&0&2&1\\0&1&1&1&1\end{bmatrix}$$

因此,该个体网的中心点“1”的中间度就是该矩阵主对角线上方三个非零数字(2, 1, 1, 1)倒数之和,即 3.5。要注意,这些计算过程都能在相关软件(如 UCINET, SPSS, SAS, STATA, R, MATLAB 等)中实现。

艾弗雷特等学者(Everett & Borgatti, 2005:35—36)利用模拟法证明了,个体中间度和整体网的中间度指标之间是高度相关的。其研究结论是,对于很多实际的数据来说,“个体的中间度是对中间度的一种很好的估计”(Everett & Borgatti, 2005:38)。

4. 结构洞指标与中间中心度指标之间的关系。

上文探讨了结构洞指标和中间中心度。在同一个网络中,应该运用哪种指标测量结构洞,这需要分析二者各自的适用范围,结合每种指标的特点并根据实际情况进行选择。不管分析个体网资料,还是整体网数据,都可采用伯特的结构

洞指标，特别是网络限制度，也可以利用中间中心度指标。但是这两种指标的适用范围有所不同，中间中心度指标的适用范围更大一些，它适用于分次有向网、2-模网、多值网、多元网，还可以分析群体的中心度(Everett & Borgatti，1999)。而结构洞指标包含四个子指标，各有其具体的含义。二者在概念、含义和适用范围等方面不同。

针对同一组关系数据，如果分别利用中间中心度和网络限制度指数进行分析，会得到怎样的结果？这两种指数之间是什么关系？这一点可以利用下面的数据分析加以说明。在管理学中有一个经典实验——霍桑实验。在这个实验过程中搜集到霍桑工厂的多方面数据，在 UCINET 中用"Wiring"代表。其中包含多个数据，对其中的一个数据，即友谊关系数据(RDPOS)进行中间中心度分析，得到的文件可命名为 FreemanBetweenness；再对该友谊关系数据进行结构洞分析，按照个体网模式分析得到的文件命名为 Holes；将得到的这两个文件合并(在 UCINET 中，点击 Data→Join 命令，按照"列"进行合并)，文件命名为 Joined；最后分析 Joined 文件各列的相似性(Tools→Similarities)，得到相关系数矩阵如表 9.5 所示，从中就可以看出这两种结构洞指数之间的关系。

表 9.5 中间中心度与结构洞指标之间的相关系数矩阵

```
SIMILARITIES
--------------------------------------------------------------------------------
Measure:            CORRELATION
Variables are:      COLUMNS
Input dataset:      H:\My Papers\DataFiles\Joined whole
Similarity matrix:  H:\My Papers\DataFiles\Joined whole-Sim
```

	Betwee	nBetwe	Degree	EffSiz	Effici	Constr	Hierar	Ego Be
	-------	-------	-------	-------	-------	-------	-------	-------
Betweenness	1.000	1.000	0.624	0.800	0.319	−0.867	−0.242	0.891
nBetweenness	1.000	1.000	0.624	0.800	0.319	−0.867	−0.242	0.891
Degree	0.624	0.624	1.000	0.928	−0.427	−0.783	−0.736	0.686
EffSize	0.800	0.800	0.928	1.000	0.279	−0.953	−0.292	0.836
Efficiency	0.319	0.319	−0.427	0.279	1.000	−0.200	0.778	0.256
Constraint	−0.867	−0.867	−0.783	−0.953	−0.200	1.000	0.283	−0.924
Hierarchy	−0.242	−0.242	−0.736	−0.292	0.778	0.283	1.000	−0.249
Ego Betweenness	0.891	0.891	0.686	0.836	0.256	−0.924	−0.249	1.000

表9.5给出了八个指标之间的相关系数矩阵。从表中可见,个体网中心度与度数之间正相关,相关系数为0.686,与有效规模之间的相关系数为0.836,与效率指数之间的相关系数为0.256,与网络限制度指数之间却是负相关,相关系数为−0.924,与等级度指数之间也是负相关,相关系数为−0.249。这些都是可以理解的。因为可以想象,一个点越是居于网络的核心,它的结构洞可能越多,所受到的网络限制度就越小,并且中间中心度越大,等级度越小。所以,中间中心度与网络限制度之间是负相关的,与等级度也是负相关的。

至于在具体分析的时候采用哪种指标,当然没有成型的规定。但是在笔者看来,由于伯特给出的结构洞指标包含四项内容,比中间中心度表达的指标更加丰富,因此笔者比较倾向于使用伯特的指数。当然,这两种指标可以对照使用。

第二节　中间人分析

在一个网络中可能存在一些中间人,在不同的网络之间也可能存在中间人。中间人往往掌握着多个群体之间的秘密,因而具有一定的竞争优势。前文已经指出,网络中存在的结构洞位置具有竞争优势。结构洞的存在界定了各种机会的大小,它与网络的密度、规模、等级度和中间性有关。在管理者的社会资本中,弱关系和结构洞占据重要地位,它与企业在市场上的表现以及市场网络等都具有重要的关联。

伯特提出了"中间人"的概念,他将中间人界定为向一个位置发送资源,却从另外一个位置那里得到资源的行动者。尽管中间人概念近年来得到了关注,但是"很少有人试图具体阐明该现象到底是什么"(Gould & Fernandez, 1989:89)。例如,古尔德和费尔南德兹(Gould & Fernandez, 1989:94—95)就认为伯特的中间人定义是有问题的,伯特没有考虑行动者所在的群体。他们认为"经纪"(brokerage)是一个中间人促进两个相互缺乏信任的行动者之间交易的过程。也可以说,只要一个行动者试图促进另外两个行动者之间的交易,我们就说他是一个掮客或者"中间人"(broker)。一般意义上的中间人在"拉关系"的时候,目的是为了获利,但是古尔德和费尔南德兹(Gould & Fernandez, 1989)所说的中间人的目的只是为了"拉关系",不管其动机是否为了获利。因此,可以简单地把中间人界定为居于中间位置之人,不管他是否能得到直接的回报。

如果一个网络可以分为相对互斥的子群体,那么行动者之间的交换关系的意义可能不同。例如,某个派别的成员可能通过敌对派别的一个成员接近敌对派别的其他成员,也可能通过自己派别的一个成员接近敌对派别的成员。在这两种情况下,两个中间人的角色显然是不同的。前者扮演类似“守门人”(gatekeeper)的角色;后者扮演“代理人”(representative)的角色。当交易网络中的行动者被分到互斥的各个子群体的时候,他们之间的关系会表现出怎样的调节结构(mediation structures)?每一种结构有着怎样的现实意义?古尔德和费尔南德兹认为,总的来说会出现如下五类中间人,并且可用统计推断来判断一个行动者的中介行为是源于交换关系的随机分布,还是来源于一种内在的趋势,即以特定方式构造这些关系的趋势(Gould & Fernandez, 1989:91—93;罗家德,2002:80—82)。可以用 B 所扮演的社会角色来命名中间人的具体角色。

一、中间人的分类

顾名思义,“中间人”指的是居于中间位置的人。具体来讲,在一个三方关系 ABC 中,如果 A 有一个指向 B 的关系,B 有一个指向 C 的关系,但是 A 没有指向 C 的关系,那么 B 就是中间人。也就是说,A 需要通过 B 才能与 C 联络上,B 因此是一个中间人。如果 A、B、C 位于不同的群体,那么中间人扮演什么角色呢?总的来说会出现如下五类中间人,并可以根据 B 扮演的社会角色来命名中间人的具体角色。

1. 协调人(coordinator):如果 b 是一个中间人,并且 a, b, c 处于同一个群体之中,这时候称 b 为协调人或者局部中介(local broker)。也就是说,协调人是在一个群体中起到中介作用的人。如果把包含多个家族的一个村落看成是一个整体网,并且假定在一个家族内部的两个家庭之间出现矛盾,该家族内部的另外一个家庭出面调解,那么该家庭扮演的角色就是“协调人”。

2. 顾问(consultant)。如果 a, c 处于同一个群体之中,而作为中间人的 b 处于另外一个群体,这时候称 b 为顾问,因为作为顾问的 b 是一个“外人”。有关此类中间人的例子就是股票中间人(stockbroker),因为经纪公司是远离其顾客的,并且在经纪公司看来,无论是买入的人还是卖出的人都一样。

3. 守门人(gatekeeper)。如果 b 是中间人,并且 b, c 处于同一个群体之中,而 a 处于另外一个群体,这时候称 b 为守门人。此时关系的发起人位于另外一个群体之中。

4. 代理人(representative)。如果 b 是一个中间人,并且 a,b 处于同一个群体之中,而 c 处于另外一个群体,这时候称 b 为代理人。

5. 联络人(liaison)。如果 b 是一个中间人,并且 a 所在的群体既不同于 b 所在的群体,也不同于 c 所在的群体,这时候称 b 为联络人。也就是说,这三个人隶属于三个群体。

在 UCINET 中,点击 Network→Ego Networks→Brokerage roles 程序,选择相应的数据即可计算出满足上述条件的五类 b 成为中间人的次数(罗家德,2002:80—82;Gould & Fernandez,1989:93)。当 b 不是 a 和 c 之间的唯一中间人的时候,可以赋予 b 一定的局部中间人取值。也就是说,如果在 a 和 c 之间存在两条长度为 2 的途径,并且其中一条经过 b,那么就可以赋予 b 的中间人指数为 1/2,而不是 1。在程序中这一项是可选的。这些中间人角色的图形如图 9.10 所示。

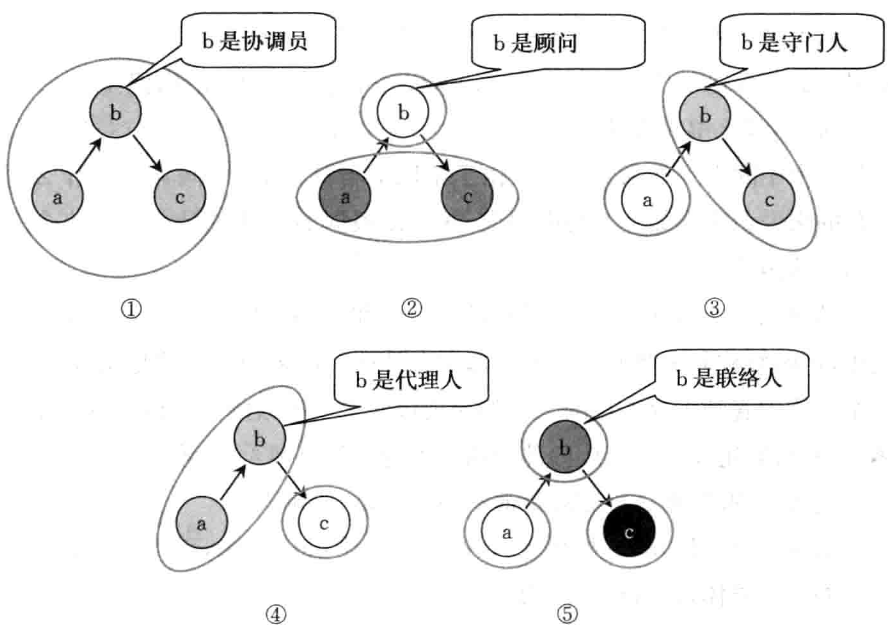


图 9.10 五类中间人

二、中间人次数的计算

如果已知一个图和各个点的分组情况,UCINET 可以针对每个点计算出五

类中间人指数，同时计算出总的中间人指数。另外，在已知群体的数量和每个群体的规模的情况下，程序还会计算出每种中间人指数的期望值。也就是说，在假设中间性与各个点的群体地位无关的情况下，程序会计算出每个点的中间指数期望值，还可以计算实际的中间指数与期望指数之比。

需要注意的是，在进行中间人分析的时候，针对的数据应该是有向图数据，并且整个网络事先的分组情况是已知的。也就是说，对于网络中的任何一个行动者来说，它隶属于哪个分组，这是在分析之前预先就知道的。正是根据这个信息和有向关系信息，才可以计算出行动者扮演着怎样的中间人角色。

分区向量(partition vector)指的是一个 UCINET 数据，它往往是一个列向量(或行向量)，其中的信息指明了每个行动者所隶属的子群体，即必须拥有一个包含着各个行动者的分区信息的 UCINET 数据。为了把数据矩阵进行分区到各个群体之中，需要指定一个向量，使之包含分区的信息。需要注意的是，程序要求研究者必须自己事先指定分区向量，也就是说，研究者自己必须首先指定网络中的行动者隶属于哪个群体，因为对上述五类中间人概念的界定都离不开群体。

在指定分区向量的时候，可以指定数据名、该向量所在的行或列和一个整数。例如，为了使用一个名为 ATTRIB 的数据的第二行，需要在点击 Network→Ego Networks→Brokerage roles 程序后出现的对话框的“partition vector”(分区向量)项目上输入“ATTRIB ROW 2”。程序会自动地读取 ATTRIB 矩阵的第二行，利用其中包含的信息来界定群体。在该行中，拥有相同值的所有行动者就被放在同一个群体之中。

以前文的“帮工第一子群”为例。由于在事前的分析中发现该群体又可分为四组，而其中的第四组只有两个村民，因此，我们将第四组和第三组合并，进而把整个数据分为三组。这三组包含的村民代号分别是：第一组：1，2，3，4，5，6，7，8，9，11，21，22；第二组：10，12，13，17，19，20，23；其余村民为第三组，具体包括：14，15，16，18，24，25。然后，在 UCINET 的数据表(spreadsheet)中，输入一行数字，赋予每个点以分区情况。也就是说，我们把点 1 分在我们指定的第一组，即在点 1 下面标记为 1；同理，点 10 属于第二个村民小组，因此在点 10 下面标记为 2。点 14 属于第三个村民小组，因此在点 14 下面标记为 3，如表 9.6 所示。

表 9.6　村民隶属村民小组情况

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1	1	1	1	1	1	1	1	2	1	2	2	3	3	3	2	3	2	2	1	1	2	3	3

经过这样的分区工作后，把该数据保存并命名为“帮工第一子群分区”。随后，在 UCINET 中，沿着 Network→Ego Networks→G&F Brokerage roles，选择“帮工第一子群分区”作为输入数据矩阵(Input dataset)，选择刚刚构建的“帮工第一子群分区”数据的第一行作为分区向量(Partition vector)(即在选择了“帮工第一子群分区”后，在后面键入 ROW 1。显然，如果用列作为分组向量，则键入 COL 1)。在具体计算过程中，Method(分析方法)一项包括“unweighted”(无权重)和“weighted”(有权重)两类。前者引导程序仅仅计算出一个点 b 处于中间人位置的次数，而不考虑有多少其他点在同样的点对 a 和 c 之间扮演同样的中间人角色。后者引导程序给出一种偏指数，该指数与扮演同样角色的其他点数成反比例关系。点击 Network→Ego Networks→G&F Brokerage roles 后，出现的对话框如图 9.11 所示。

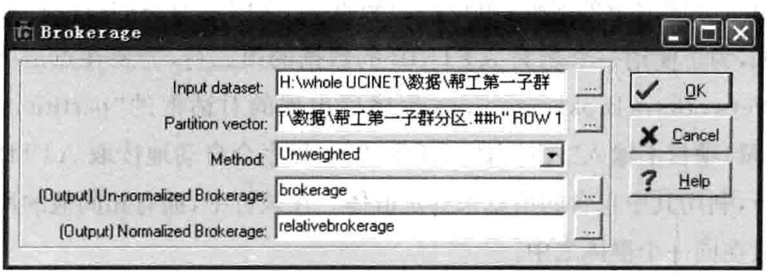


图 9.11　中间人分析对话框

点击 OK 后得到如表 9.7 所示结果。

由表 9.7 可见，点 1 在整个三组中扮演的“协调员”的次数是 17 次，扮演“代理人”的次数是 25 次，扮演“联络人”5 次。点 1 扮演五种中间人共 83 次。对其他点扮演中间人角色的分析与此类似。这些数值表明了点 1 在不同小组中扮演不同类中间人的情况。表 9.8 给出了 25 个点(代表 25 个村民家庭)在三组中扮演的中间人的分类情况。

表 9.7　五类中间人指数

GOULD & FERNANDEZ BROKERAGE MEASURES

Input dataset:　　H:\whole UCINET\数据\帮工第一子群
Partition vector:　"H:\whole UCINET\数据\帮工第一子群分区.##h" ROW 1
Method:　　UNWEIGHTED
Number of classes:　3

Un-normalized Brokerage Scores

		1 Coordinat	2 Gatekeepe	3 Represent	4 Consultan	5 Liaison	6 Total
1	1	17	34	25	2	5	83
2	2	14	35	23	2	8	82
中间略去第 3 到第 8,第 9, 22, 11 这 9 个点的中间人指标							
21	21	1	0	0	0	0	1
13	13	0	0	0	0	0	0
中间略去第 20, 12, 10, 17, 19 这 5 个点的中间人指标							
23	23	1	3	1	1	0	6
18	18	6	22	14	2	5	49
中间略去第 15, 16, 14, 24 这 4 个点的中间人指标							
25	25	2	17	13	5	10	47

表 9.8　每个点在三组中扮演的中间人指数

Group to Group Brokering for each Node

Node 1 (group 1)

	1	2	3
1	17	19	6
2	22	2	3
3	12	2	0

………(中间略去点 2 到点 24 的中间人情况)

Node 25 (group 3)

	1	2	3
1	5	7	14
2	3	0	3
3	9	4	2

表 9.8 给出了每个点在各个小组之间扮演中间人的次数情况。可以看出,在点 1 所扮演的 82 个中间人中,属于第 1 组内的中间人有 17 个,属于从第 1 组指向第 2 组的中间人有 19 个,属于从第 2 组指向第 1 组的中间人有 22 个,属于从第 2 组指向第 3 组的中间人有 3 个,等等。对其他点的分析与此类似,此处不再赘述。

在计算结果中还给出每个点的相对中间人次数(实际次数除以期望次数)和期望次数(在已知组数和每组的规模条件下的期望次数),这里从略。

总之,进行中间人分析的目的是为了分析在整个网络中,不同的行动者在不同的组中扮演怎样的中间人角色,这样可以深入了解行动者的关系属性特征。

第三节　其他个体网指标的计算

UCINET 针对个体网分析的内容还有很多,下面简要介绍。

一、个体网的指标

可以计算的个体网指标(Egonet basic measures)包括个体网的规模(Size)、关系总数(Ties)、最大可能的点对数(Pairs)、密度(Density)、平均距离(AvgDist)、直径(Diameter)、弱成分的数量(nWeakComp)、在 2-步内可达的点数与个体网规模之比(2StepReach)、可达的效率(ReachEffic),即在 2-步内可达的点数除以个体网的各个成员的个体网规模总和之比、中间人(Broker),即个体位于两个点之间的捷径的次数。要计算个体网的各种指标,需要点击 Network→Ego Networks→Egonet basic measures,选择相关的数据即可计算出结果。

例如,如果计算一个名为"紧密关系表"的整体网络数据中各个行动者(这里指的是村民)的"个体网络密度",沿着上述路径得到的对话框如图 9.12 所示。

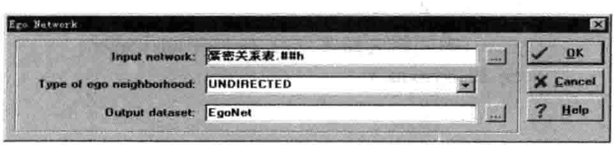


图 9.12　个体网指标分析对话框

点击 OK 后得到的计算结果如图 9.13 所示。

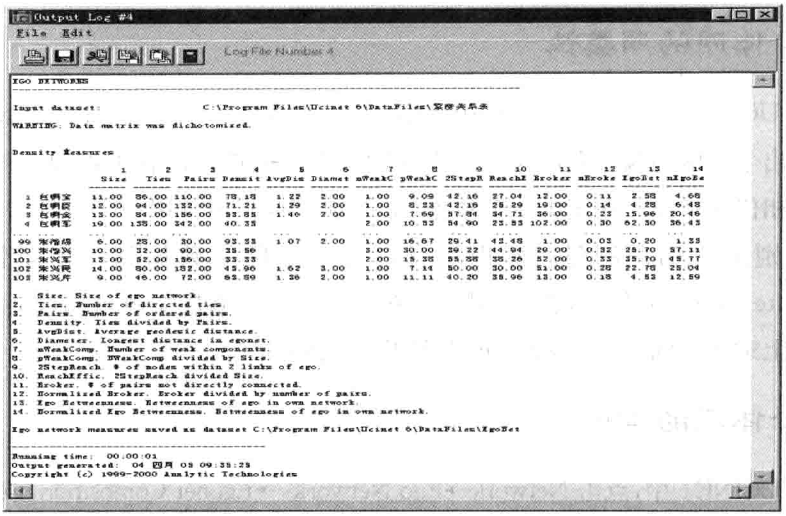


图 9.13　个体网指标分析结果

在这个结果中,已经对各个个体网测度作了说明。例如,“Size”是个体网的“规模”(不包括“自我”)。显然,第一个村民的个体网规模为“11”,说明与他有“紧密关系”的村民有 11 家。需要解释的其他测度有:“Ties”一项是“关系总数”,指的是“个体网成员之间的关系总数,不包括各个成员与‘自我’之间的关系”;“Pairs”是个体网中各个成员之间在理论上“所有可能存在的关系总和”,该值也等于(网络规模数)×(网络规模数－1)。例如第一个村民的个体网规模为“11”,那么该行动者的个体网在理论上可能存在的各种关系总数“Pairs”等于 11 × 10 ＝110 个。因此,个体网的“密度”,即“Densit”一项(即第四个输出的个体网测度)就是“实际存在的关系总数”除以“理论上可能存在的最多关系总数”,即相当于“Ties”一项的值除以“Pairs”一项的值得到的百分比。对于第一个行动者来说,该个体网的密度值为 86/110 ＝ 78.18%。其他密度值的计算与此类似。“Avgdist”是“平均距离”,是图论意义上的各个个体网成员之间的平均捷径距离,这个值只针对每个点都可达的网络来计算。“Diameter”是“直径”,指的是个体网中的最长捷径距离(geodesic distance),无限长的距离不作标记。“Nweak-Comp”是“个体网中的弱成分总数”。“PweakComp”是“弱成分数目占网络规模的比例”。“2StepReach”是“自我”点在 2 步内可达的成员总数。“ReachEffic”是“可达效率”,具体指的是“2 步内可达的点数”除以每个成员的个体网规模的总数。

二、个体网的同类性

在 UCINET 中，点击 Network→Ego Networks→Egonet Homophily 程序，会计算出个体网的核心点就某项特定属性而言与其他点的各种相似性测度，它是一种相似性指标。在定类变量数据(如性别)中，相似性用两个行动者隶属于同一性别的比例来测量。该程序会针对一个网络数据和一个属性数据(attribute dataset)进行分析，测量的是核心点的个体网成员就该属性而言在多大程度上隶属于同类，当然要根据很多指标来测量。

三、个体网的构成

在 UCINET 中，点击 Network→Ego Networks→Egonet Comosition，该程序可以提供核心点的网络成员就某种连续变量(或离散变量)方面值的描述统计量。

四、诚实中间人指数

在 UCINET 中，诚实中间人指数(Honest broker index)程序会针对每个点计算该点在网络中成为中间人的次数，以及该点所调节的点对数等指标。例如，针对上述“帮工第一子群”来说，在 UCINET 中，沿着 Network→Ego Networks→Honest broker index，选择帮工第一子群数据，计算结果如表 9.9 所示。

表 9.9　诚实中间人指数

HONEST BROKER INDEX

Input Network:　　H:\whole UCINET\数据\帮工第一子群
Ego must trust alters?　　YES
Output dataset:　　H:\whole UCINET\数据\帮工第一子群-HBI

Honest Broker Indices

	Size	Pairs	HBI0	HBI1	HBI2	nHBI0	nHBI1	nHBI2
	------	------	------	------	------	------	------	------
1	14.000	91.000	17.000	20.000	54.000	0.187	0.220	0.593
2	12.000	66.000	14.000	7.000	45.000	0.212	0.106	0.682
3	12.000	66.000	16.000	7.000	43.000	0.242	0.106	0.652
……(第 4 行—第 23 行略)								
24	1.000	0.000						
25	8.000	28.000	4.000	10.000	14.000	0.143	0.357	0.500

这里所界定的个体网仅仅考虑点入度。HBI0 代表纯中介(pure brokerage),即在中间人所联络的任何两个人之间不存在关系。HBI1 代表弱中介(weak brokerage),即在中间人所联络的任何两个人之间允许存在一条有向关系。HBI2 代表非中介(non-brokerage),即在中间人所联络的任何两个人之间存在双向关系。可见,点 1 作为中介的次数是 14 次,共联络了 91 对关系,其中 17 对关系属于 HBI0 类型的,20 对关系属于 HBI1 类,54 对关系属于 HBI2 类的。这说明点 1 所联系的大多数人之间都存在双向关系,或者说,点 1 所起到的中间人的作用不太大,因为他所联系的人之间大多数都存在双向联系。对其他点的"诚实中间人指数"分析的结果也作类似解释。

对整体网中的个体网的分析还有很多内容,请参见 UCINET 中的 Network→Etras→Change Models→ICT-Indidicual Change Tendencies 和 Network→Ego Networks 中的其他选项,特别是其中的 Longitudinal 中的四个选项则给出了分析个体网演变的技术路径。感兴趣的读者可尝试利用这些程序计算整体网中的个体网的诸多指标,这里从略。

本 章 小 结

本章介绍了结构洞的含义,它包含凝聚性和结构对等这两种界定方式。结构洞成为社会资本的机制在于,跨群体的中间人会更熟悉另类想法和行为,因而给自己带来更多的观念选择的机会,这种视野优势是结构洞成为社会资本的机制。本章还介绍了结构洞指数的两种测量方法:伯特的结构洞指数和中间中心度指数,并对二者之间的关系进行了比较。第二部分对中间人的五种类型进行了分析,并且展示其具体运算。本章同时对个体网分析的其他指标(同类性、构成性、诚实指数等)进行了简要介绍。

第十章

一致性分析

假设有很多人参加一项重要考试，在阅卷的时候教师把单选题的标准答案丢了（当然这是不太可能的！），又假设阅卷老师无法确定标准答案是什么。在这种情况下，一致性分析就派上了用场：它能帮助我们找到标准答案！

第一节　一致性分析的含义

一致性分析（consensus analysis）既是一种理论，又是一种方法（本章主要参考 Borgatti, 1997）。作为一种理论，它研究的问题是，多个个体在“测验”问题的回答方面有多大一致性，进而表明他们在多大程度上拥有知识。作为一种方法，可用它来揭示在面对某些类型的文化内部变动性（certain kinds of intra-cultural variability）的条件下，对一系列问题在文化上给出的正确回答。与此同时，研究者可用它来评价一个线人拥有的关于某种固定文化领域的知识的范围。

假设在“社会学导论”课程考试中有这样一道单选题：“《自杀论》的作者是________”，随后给出五个选项。如果将学生的选择看成数据，那么我们就会得到一个长方形的答题者—问题矩阵 X（respondent-by-question matrix），其中 x_{ij} 表示学生 i 在第 j 个问题上的选项。矩阵的每一行都是一个由数字构成的向量，其取值范围从 1 到 5，代表每个固定的学生对每个问题的回答。标准答案也是一个向量，其取值范围相同，代表教师给出的回答项。按照标准答案对学生的答题进行评判，可以得到学生考试的成绩。实际上这个过程就相当于计算学生向量与教师向量之间的相关关系。如果这两个向量在全部问题上都相同，那么学生就得到满分。如果不同，就表明学生的回答有误，不能得到满分。这里便引出一个重要问题：学生的考试成绩实际上是一个测度，它测量的是学生向量与教师向量之间的相似性。这种相似性测度就是简单匹配系数（simple match coefficient），即学生和教师给出的一致答案的问题数除以问题总数。

推而广之，我们也可以计算任何两个匹配向量之间的相似性。例如，可以计

算两个学生向量之间的相似性。如果两个学生都得满分，那么他们的向量将相同(identical)，当然前提假设是每个问题的答案只有一个。如果两个学生回答问题都很好，我们就可以期望他们的答案有很多相似性。反之，他们的回答越是错误，二者答题向量的相似性也就越小。如果两个学生的回答都很少有正确的，那么两个答题向量之间的相似度将小于随机水平。

在这种思想实验的指引下，我们把学生之间的一致性(即回答项之间的相似性)看成是一个函数，即涉及每个学生关于所涉及知识的一个函数。下面阐述该函数的条件。

首先，我们假定要回答的问题是单选题，即针对每个问题的正确答案有且只有一个。我们同时假设学生 i 的回答具有如下模式：

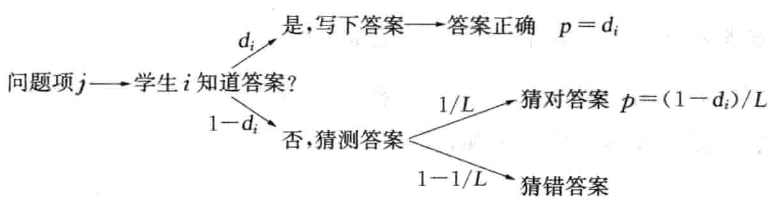


图 10.1 回答问题对错的概率

如果学生知道答案，那么他的回答就是正确的。如果学生不知道问题的答案，他就会随机地猜测。令 d_i 表示第 i 个学生知道任何已知问题答案的概率。可以将 d_i 看成是第 i 个学生知道有关一个既定问题的全部可能答案的比例，那么学生 i 不知道问题答案的概率就是 $1-d_i$。如果他不知道一个既定问题的答案，那么他可以猜测，猜对答案的概率是 $1/L$(其中 L 是可选项的数目，本例中是“5”)。我们当然假定他不能排除一些明显错误的答案选项。如果可以排除一些错误的选项，那么猜对的概率就是 $1/L^*$(其中 L^* 是去掉一些明显错误的答案后剩下的选项数)。不管在什么情况下，正确地得到一个问题的答案的概率等于知道正确答案的概率，再加上猜对答案的概率，即得到第 j 个问题的正确答案的概率为 $m_i=d_i+(1-d_i)/L$。

从这个公式中可以解出知道正确答案的概率值 d_i 等于

$$d_i=(Lm_i-1)/(L-1)$$

利用这个简单的模型，我们可以对上述思想实验进行细化，得到一个更加精

细的陈述。首先,令两个学生 i 和 j 分别以认知水平 d_i 和 d_j 给出针对同一个问题的相同答案。这里可能出现四种情况:

1. i 和 j 都知道正确答案,即

$$p(\text{都知道答案})=d_id_j$$

2. 学生 i 知道正确答案,学生 j 猜对答案,即

$$p(i\ \text{知道},j\ \text{猜对})=d_i(1-d_j)/L$$

3. 学生 j 知道正确答案,学生 i 猜对答案,即

$$p(j\ \text{知道},i\ \text{猜对})=d_j(1-d_i)/L$$

4. 二者都不知道正确答案,但是都猜同一个答案(不管其对错),即

$$p(\text{虽都不知道正确答案,但都猜的是同一个})=(1-d_i)(1-d_j)/L$$

针对任何一个问题来讲,如果令 i 和 j 都猜到同一个答案的概率为 m_{ij} 的话,那么它就是上述四个概率之和,即

$$\begin{aligned}m_{ij}&=d_id_j+d_i(1-d_j)/L+d_j(1-d_i)/L+(1-d_i)(1-d_j)/L\\&=d_id_j+(1-d_id_j)\ /L\end{aligned}$$

如此看来,i 和 j 之间的一致性就是 m_{ij}。这便得出一个重要的理论结果:在给定上述测验条件和学生回答模型的基础上,一致性就意味着知道有关问题的知识,反之亦然。有关这种模型的一些前提性的假设可汇总如下:

1. 单选题假设,即每个问题有且只有一个正确答案。

2. 局部依赖性假设(local independence),即学生们的回答是独立的。如果学生不知道答案,那么他将随机从答案中选取。

3. 项目同质性假定(item homogeneity),即各个问题是从一系列可能的问题中选出来的。因此,学生 i 知道一个问题答案的概率 d_i 对所有问题来说都一样。因此,所有的问题都是关于同一个主题的问题,就此而言,某个特定的学生拥有的知识水平是一定的。这一假设体现在回答模型中,就是利用单个参数 d_i 来刻画一个回答者知道答案的概率。

现在,我们看关键结果。在上文最后一个等式中,m_{ij}(学生 i 和 j 给出同样回答问题的比例)是已知的。我们可以根据两个学生的回答向量(response vectors),计算二者匹配的比例。在一次测验中必然拥有标准答案,可以计算 d 个

参数，因为它们仅仅是正确回答问题的比例减去随机猜对的比例。但是假设我们没有标准答案，等式便可以写成如下：

$$d_i d_j = (Lm_{ij} - 1)/(L - 1) = m_{ij}^*$$

该等式意味着，未知的两个参数 d 之积等于在测验中观察到的两个学生在回答问题方面的相似性，并且根据猜测进行了修正。观察到的相似性是各个 m_{ij}。在根据随机因素进行修正之后，我们得到了 m_{ij}^*，它恰好是对观察到的相似性的重新计量(rescaling)。这个新等式可利用最小剩余因子分析(minimum residual factor analysis)进行求解，从而得到对 d 个参数的最小二乘估计值(least squares estimates of the d parameters)。

换句话说，即便没有标准答案，我们仍然完全可以通过对学生—学生相似性的因子分析来考察每个学生掌握多少知识，并且可以判断谁知道答案，谁不知道答案。我们也可以判断每个问题的答案必定是哪个。也就是说，我们可以在众多已知答题者如何选择答案项的情况下，确定一个既定问题的答案最有可能是什么。例如，如果 20 个学生对全部问题的 90%的回答都是正确的，而学生们认为第 7 个问题的答案是“b”，那么不管大多数学生是怎样回答的，该问题的答案不是“b”的概率都将极低。

对于常规的考试来说，这个结果的意义当然不大。但是，普通人想不到的是，这个结果对于文化人类学家来说却极端重要，因为他们一般不知道自己所研究的问题的答案是什么，需要找到答案。而在探究答案的时候，人类学家面临着文化多样性的存在(cultural variability)，如何在多样性中寻找比较“一致”的答案，这是一个需要研究的问题。如果我们向线人样本(a sample of informants)询问某些基本问题，可能得到一系列矛盾的回答。这里我们不讨论个人的选项，只关注一般意义上的问题，即所有回答者都一致同意有一个“正确”答案的问题，但是该问题是什么，却没有达成共识。有时候，这种不一致性是由亚文化的多样性(subcultural variability)造成的：真理的存在如同亚文化的存在一样是多样的。即便在一种亚文化中，知识也可能是多样的，导致的结论也有多个。因此，一致性分析的方法论允许人类学家(1)找到正确答案，并且(2)判断谁知道一个既定问题的答案，谁不知道答案。

有一点很重要需要指出来，即一个问题的“正确答案”是在文化意义上界定的概念。我们所讨论的“真”并不是自然科学意义上的“真理”。知道“地球是平

的吗?”这个问题的正确答案与理解天文学或者地理学是无关的,它是一个人对某个既定群体的文化进行把握的函数。

一致性分析的方法论依赖于上述三个假定。如果将其置于人类学语境中,可表述如下:

1. 一种文化(one culture)。我们假定,不管文化现实是什么,它对于每个人来说都一样。就对一个已知问题的回答而言,不存在有系统差异的多种亚文化。所有的变异性(variability)都是由知识的多样性(variations in amount of knowledge)带来的。

2. 独立性。激发人们给出某种既定答案的唯一动力就是从文化的角度给出正确的答案。当线人们不知道答案的时候,他们会独立选择答案。换句话说,要对每个线人进行单独访谈,不能进行群体访谈,并且尽量阻止回答者进入到一个“答案集合”(response set),例如要尽量阻止回答者总回答“是的”。

3. 一个领域(one domain)。所有的问题都来自同一个问题域:你绝不能将关于网球的问题与关于星球的问题混在一起,因为一个人关于网球的认知可能完全不同于对星球的认知。

如果这些假设成立,你就可以相信自己对一个线人拥有知识的程度的判断是对的,也可以相信自己对正确答案的判断。另外,有关一致性分析的软件ANTHROPAC可以检验这些假设是否成立,或者更精确地说,软件可以检验这些假设是否不成立。一种检验方法便是计算 M^* 矩阵的特征值(eigenvalues)。如果计算结果表明,大的特征值多于一个,那么关于单一文化的假设(the one culture assumption)就不成立。例如,如果在计算结果中出现两个大的特征值,则表明至少存在两个真理(即存在两种完全不同的回答模式)主导着线人的回答。该程序会计算出第一个特征值与第二个特征值之比。我们在判断的时候可以坚持一个大致的原则:如果该比例小于3,那么单一文化假设就没有得到支持;如果比例达到10∶1,那么该假设就得到较强的支持,但是永远不能证明其有效性。

第二节　一致性分析举例

假设在一次考试中,我们得到了10个考生在前四道单选题(分别是q1,q2,q3,q4)上的答题情况,每个单选题都有四个选项1,2,3,4。假设这10位同学

在四个问题上的回答情况如表 10.1 的 10 行 4 列数据所示。

表 10.1　10 位同学回答 4 个问题的答案

学生代码	q1(问题 1)	q2(问题 2)	q3(问题 3)	q4(问题 4)
1	1	3	4	2
2	2	2	4	2
3	1	3	4	2
4	1	2	4	3
5	1	3	3	2
6	3	4	4	2
7	1	3	3	1
8	3	3	4	2
9	4	3	2	2
10	1	3	4	2

例如，第 5 名同学在 q3 问题上给出的答案是“3”。将这些数据输入 UCINET，命名为“consensus analysis”，并保存在硬盘上(记住保存在硬盘上的具体路径)，然后在 UCINET 中，沿着 Tools→Consensus Analysis 这条路径，在硬盘中选出刚刚输入的文件 consensus analysis，如图 10.2 所示。

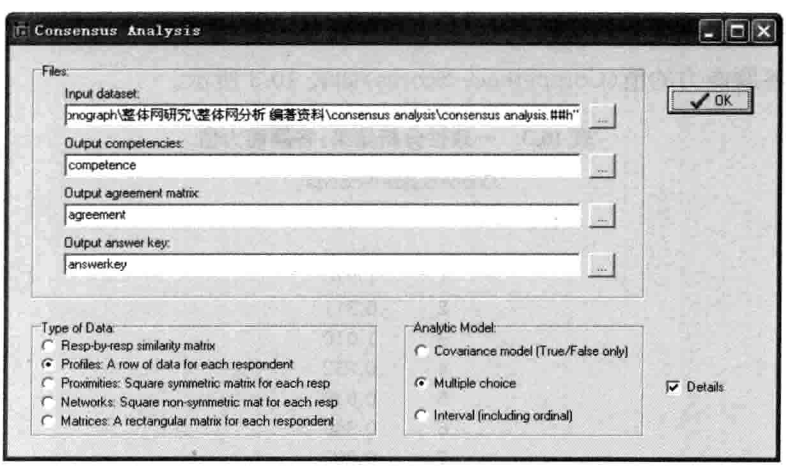


图 10.2　一致性分析对话框

点击 OK，即可得到计算结果。

表 10.2　一致性分析结果

```
CONSENSUS ANALYSIS
--------------------------------------------------------------------------------
Type of data:      Profiles: A row of data for each respondent
Analytic model:    Multiple choice

Agreement among respondents
```

	1	2	3	4	5	6	7	8	9	10
	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1	1.00	0.33	1.00	0.33	0.67	0.33	0.33	0.67	0.33	1.00
2	0.33	1.00	0.33	0.33	0.00	0.33	−0.33	0.33	0.00	0.33
3	1.00	0.33	1.00	0.33	0.67	0.33	0.33	0.67	0.33	1.00
4	0.33	0.33	0.33	1.00	0.00	0.00	0.00	0.00	−0.33	0.33
5	0.67	0.00	0.67	0.00	1.00	0.00	0.67	0.33	0.33	0.67
6	0.33	0.33	0.33	0.00	0.00	1.00	−0.33	0.67	0.00	0.33
7	0.33	−0.33	0.33	0.00	0.67	−0.33	1.00	0.00	0.00	0.33
8	0.67	0.33	0.67	0.00	0.33	0.67	0.00	1.00	0.33	0.67
9	0.33	0.00	0.33	−0.33	0.33	0.00	0.00	0.33	1.00	0.33
10	1.00	0.33	1.00	0.33	0.67	0.33	0.33	0.67	0.33	1.00

No. of negative competencies:	0
Largest eigenvalue:	4.385
2nd largest eigenvalue:	1.022
Ratio of largest to next:	4.291

特征根之比为4.291，大于3，表明存在着单一的回答模式。计算出来的各个学生答题能力的值(Competence Scores)如表10.3所示。

表 10.3　一致性分析结果：答题能力值

Competence Scores:

	1

1	1.010
2	0.311
3	1.010
4	0.232
5	0.639
6	0.346
7	0.283
8	0.682
9	0.321
10	1.010

该值越大，表明学生正确回答问题的能力越强。可见，第 1、第 3 和第 10 号学生正确回答问题的能力最强，其次是第 8、第 5 号学生等。根据数据分析得到的每个问题的答案如表 10.4 所示。

表 10.4　一致性分析结果：准确答案

Answer Key

	Page
	----
q1	1.00
q2	3.00
q3	4.00
q4	2.00

可见，第 1，第 2，第 3，第 4 个问题的答案分别是 1，3，4，2。

最后需要补充的是，一致性分析不仅仅适用于分析上述类型的数据，它最初来源于认知人类学，是为了分析结构化的访谈数据（structured interview data）而发展出来的一种技术。在文化人类学或认知人类学中，一致性分析可以产生三项有意义的结果：(1)不同的线人在关于某种知识、信念或者事件方面的一致性程度指标；(2)根据各个人的回答，得到在“文化上正确的”信息；(3)给出每个人的指标，用来表征此人对该领域知识的掌握程度。

一致性分析并不仅仅适用于高度一致性的领域，也适用于弱一致领域、易动性领域（turbulent domains）和多中心领域（它又包括两类，一类是亚文化领域，其中包含两个或多个中心，这些中心虽然不同，但是并不互斥；第二类是竞争领域，即在同一人群中人与人之间存在相反的见解）（Caulkins，Susan & Hyatt，1999：7）。考尔金斯、苏珊和海厄特（Caulkins，Susan & Hyatt，1999）探讨了在“低一致性领域”如何对多样性进行概念化的分类，具体包括(1)弱一致性（weak agreement）；(2)干扰项（turbulent）；(3)亚文化（subcultural）；(4)竞争的领域。这种分类有助于测量和解释组织和社会运动中的多样性和变迁。

感兴趣的读者可参见这篇文章，可对涉及文化领域中的问题进行一致性分析。

本 章 小 结

一致性分析来源于认知人类学，是为了分析访谈数据而发展出来的一种定量技术，其目标是测量多个个体(线人、被调查者)在回答问题方面有多大一致性。它能够揭示文化的内部变动性。在文化人类学领域，一致性分析的结果有：不同的线人关于某种知识的一致性指标；得到在"文化上正确的"信息；测量出每个人对该领域知识的掌握程度。一致性分析也适用于弱一致领域、易动性领域和多中心领域，其适用范围比较广。

第十一章

2-模网研究

人们往往是通过他们所隶属的组织或者群体而联系在一起的。个人与组织之间的网络是 2-模网络(2-mode networks),个体与组织或群体之间因而存在“二元性”。在社会网络分析中,所谓“二元性”(duality)的含义是,群体之间的关系就意味着个体之间的关系。如何分析二元性,如何研究 2-模网络?本章将介绍 2-模网络的一些分析方法。总的来说,对 2-模网络可以进行图形分析、二部数据结构分析、2-模网络的定量分析和 2-模网络的分派分析等。下面将分别加以介绍。

第一节　2-模数据的矩阵分析和图形分析

在多数情况下,关系数据都是微观层次的数据,但是研究人员却用一些技术来推断宏观层次的社会结构。例如,我们可以利用个体层次的关系推导出诸如派系、n-派系、k-丛这样的宏观结构。但是,本书第一章已经指出,有一类数据描述的是两类群体,在特殊情况下描述一群行动者和诸多事件之间的关系,这就是 2-模网数据,也称之为隶属关系数据。2-模网络数据为我们分析“微观和宏观”、“个人和集体”、“主观和客观”关系提供了可能。例如,20 世纪 40 年代,美国学者戴维斯(Davis)及其同事收集到 18 位妇女参与 14 个事件的信息。在该数据中,如果考察每位妇女参与了哪些事件,就可以解释这个 2-模网络的结构,找到其中存在的派系和分组情况,进而赋予每个事件以一定的意义。反之,我们也可以分析作为宏观层次结构的各个事件如何影响到每位妇女的选择,同时可以看到各个事件之间的相似性。

正是利用这个例子,布雷格(Breiger, 1974)探讨了“个人和群体之间的二元性”(the duality of persons and groups)问题。布雷格关注的二元论问题是指如下情形,即每个人是如何通过他们所隶属的机构而形成社会结构的;反过来,社会结构又是如何通过一种制度化的安排限制并形塑着其中个体的行为的。

诸如此类的 2-模网络(或者微观—宏观结构)很多。推而广之,从更高层次

上讲,社会世界也是一类"嵌套"型的结构,其中的个体(和大型的结构)都嵌套在更大的结构之中。社会学分析的核心内容之一就是诸如"结构和能动"(structure and agency)、"个体和集体"或者"微观和宏观"(macro and micro)之间的张力问题。以往对此类问题的量化研究不多,现在,这类问题可以利用 2-模网络技术加以分析。

为了分析的方便,下面主要利用一种假想的 2-模网络数据,即 14 名同学参与 7 个学习小组的数据。该数据具有两模:学生和学习小组。它是一种二值数据,表明了每个同学参与学生小组的情况。

一、2-模网络向 1-模矩阵的转化

2-模数据一般用长方形矩阵来表示,可将它转换为两个 1-模数据,也可以转换为一个 2-部 1-模矩阵,在此基础上可画图分析。我们假想的学生—学习小组数据可用长方形数据矩阵来保存。首先看看该数据的具体情况,如表 11.1 所示。

表 11.1　14 名同学参与 7 个小组学习的情况

	maths	politics	english	geography	chemistry	speech	biology
A	1	0	0	1	0	0	0
B	0	1	1	0	0	1	0
C	1	0	1	1	0	0	1
D	0	1	0	0	0	1	0
E	0	0	1	0	1	0	0
F	1	0	0	1	0	0	1
G	0	1	0	0	1	1	0
H	0	0	0	1	1	0	0
I	1	0	0	0	0	1	1
J	0	1	1	0	0	1	0
K	0	1	0	1	0	1	0
L	1	0	0	1	0	0	1
M	1	0	1	0	0	1	1
N	0	1	1	0	1	1	0

我们把该数据命名为"affiliation.14.7",并保存在 H 盘中。从数据中可以看到,第一个同学(即 A)参加了"数学"、"地理"小组的学习;第 12 个同学(即 L)同学参加了"数学"、"地理"和"生物"小组的学习。其他同学的参与情况从表中可一目了然。

（一）数据的转换——从 2-模到 1-模

分析 2-模数据的最常见的方法是把该数据转换为两个 1-模数据，进而考察每一类点之间的关系。例如，我们可以创建一种“学生—学生关系”数据，它测量的是每一对同学共同参与同一学习小组的次数。我们也可以创建一种事件—事件（小组—小组）关系数据，它记录的是每一对事件（学习小组）之间共享的行动者（学生）的次数。在 UCINET 中，点击 Data→Affiliations（2-Mode to 1-Mode）工具就可以实现这种想法。点击之后出现对话框，在 H 盘中选出我们要分析的数据“affiliation.14.7”，如图 11.1 所示。

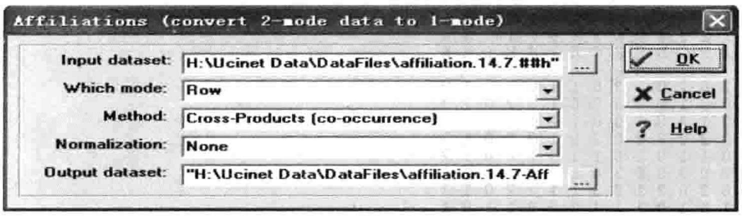


图 11.1　2-模网向 1-模网转化对话框（行模式）

在具体转换的时候可有多种选择：

1. 就模态的选择来说，有两种选项。如果选择行模式（row mode），生成的则是以行动者为计算标准的矩阵，即行动者—行动者关系矩阵；如果希望创建一种事件—事件 1-模矩阵，则需要选择“列模式”（column mode）。

2. 转换的方法有两种。

（1）对应乘积法（cross-product method），该方法考虑到行动者 A 所在行的每一项，分别乘以行动者 B 的对应项，然后加总。通常情况下，这种方法用于二值数据，因为乘积是对“共同发生”次数的累加。对于二值数据来说，只有当每个行动者在某个事件上都出现的时候，乘积的值才是 1。因此，在各个事件上的总和就等于行动者共同参与的事件的次数，这同时也测量了关系的强度。

对于本例来说，如果两个行动者都没有参与事件，结果为 $0\times0=0$，如果一个行动者参与事件，另外一个人不参与，结果是一样的：$0\times1=0$。只有当两位同学都参与了一个事件的时候，即 $1\times1=1$，在结果矩阵中二者对应之值才是 1。

（2）最小值方法（minimums method），该方法考察在每个事件上的两个行动者的每一项，并选择出最小值。对于二值数据来说，计算的结果与差积计算方法（cross-product method）是一样的。对于多值数据来说，最小值方法的实质就

是:两个行动者之间的关系等于二者与事件之间关系的最小值。当初始数据是多值数据的时候,这种方法是常用的。

如果选择“行”模式,并且利用“对应乘积法”,得到的转换矩阵就是同学—同学关系矩阵。如表 11.2 所示。

表 11.2 利用 Data→Affiliations 程序创建的同学—同学关系

AFFILIATIONS

Input dataset: "H:\Ucinet Data\DataFiles\affiliation.14.7
Dimension: ROWS
Method: Cross-Products (co-occurrence)
Normalization: None

	1 A	2 B	3 C	4 D	5 E	6 F	7 G	8 H	9 I	10 J	11 K	12 L	13 M	14 N
1 A	2	0	2	0	0	2	0	1	1	0	1	2	1	0
2 B	0	3	1	2	1	0	2	0	1	3	2	0	2	3
3 C	2	1	4	0	1	3	0	1	2	1	1	3	3	1
4 D	0	2	0	2	0	0	2	0	1	2	2	0	1	2
5 E	0	1	1	0	2	0	1	1	0	1	0	0	1	2
6 F	2	0	3	0	0	3	0	1	2	0	1	3	2	0
7 G	0	2	0	2	1	0	3	1	1	2	2	0	1	3
8 H	1	0	1	0	1	1	1	2	0	0	1	1	0	1
9 I	1	1	2	1	0	2	1	0	3	1	1	2	3	1
10 J	0	3	1	2	1	0	2	0	1	3	2	0	2	3
11 K	1	2	1	2	0	1	2	1	1	2	3	1	1	2
12 L	2	0	3	0	0	3	0	1	2	0	1	3	2	0
13 M	1	2	3	1	1	2	1	0	3	2	1	2	4	2
14 N	0	3	1	2	2	0	3	1	1	3	2	0	2	4

1-mode matrix saved as dataset "H:\Ucinet Data\DataFiles\affiliation.14

对角线上的值代表单个行动者参与事件的数目。可见,第 1 行第 1 列的值为 2,说明 A 同学参与了两个小组(该行乘以自身的对应项得到的值)。第 1 行和第 12 列的值也是 2,表明 A 同学和 L 同学共同参与了两个学习小组的学习。对其他值的解释与此类似。当然,利用这种算法可以揭示出一些信息,但是不能揭示出全部信息。至少两个同学共同参与的学习小组到底有哪些?这样的信息必须结合原始数据才能说明。

如果希望创建小组—小组 1-模矩阵,则需要选择“列模式”(column),如图 11.2 所示。

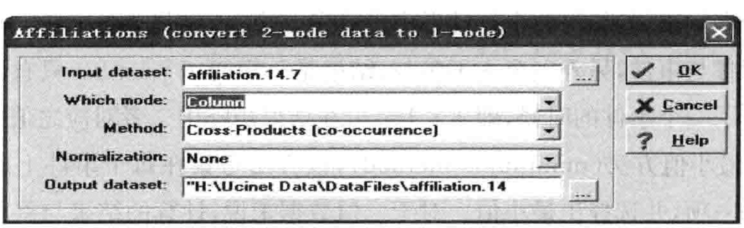


图 11.2 2-模网向 1-模网转化对话框(列模式)

点击 OK 后得到如表 11.3 所示的结果。

表 11.3　小组共享成员

```
AFFILIATIONS
--------------------------------------------------------------------------------

Input dataset:                    "H:\Ucinet Data\DataFiles\affiliation.14.7
Dimension:                        COLUMNS
Method:                           Cross-Products (co-occurrence)
Normalization:                    None

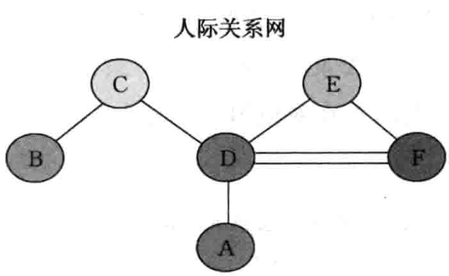

                  1 2 3 4 5 6 7
                  m p e g c s b
                  - - - - - - -
  1       maths   6 0 2 4 0 2 5
  2    politics   0 6 3 1 2 6 0
  3     english   2 3 6 1 2 4 2
  4   geography   4 1 1 6 1 1 3
  5   chemistry   0 2 2 1 4 2 0
  6      speech   2 6 4 1 2 8 2
  7     biology   5 0 2 3 0 2 5

1-mode matrix saved as dataset "H:\Ucinet Data\DataFiles\affiliation.14
```

可见，"数学"小组共有 6 个同学参加(即第一行第一列的值为 6)。有两位同学既参加了数学小组，又参加了英语小组(即第一行第三列的值为 2)。对其他值的解释与此类似。当然，如果追问哪些同学参加了什么小组，这还要结合原始数据才能说明。作为结果的行动者—行动者矩阵或者事件—事件矩阵都是 1-模多值矩阵，表达了关系的强度。在此基础上，任何用来分析 1-模网络的方法(如中心性分析等)都可用来分析这种矩阵。

(二) 数据从 2-模到 1-模转换的矩阵算法

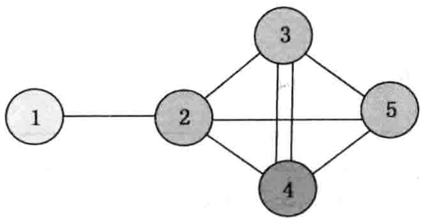
为了更清楚地展示数据转换的过程，下面用一个简单的例子详细说明(本例参见汉尼曼的讲义和穆迪的讲义)。假设现在有一个数据表达了 6 人(A, B, C, D, E, F)参加 5 个事件的情况。



	A	B	C	D	E	F
A	0	0	0	1	0	0
B	0	0	1	0	0	0
C	0	1	0	1	0	0
D	1	0	1	0	1	2
E	0	0	0	1	0	1
F	0	0	0	2	1	0

图 11.3　6 点图及其矩阵

可见，A 和 D 共同参加 1 个事件，D 和 F 共同参加 2 个事件，其他情况如图 11.3 所示。5 个事件之间共享的成员情况如图 11.4 所示。



	1	2	3	4	5
1	0	1	0	0	0
2	1	0	1	1	1
3	0	1	0	2	1
4	0	1	2	0	1
5	0	1	1	1	0

图 11.4　5 个事件的关系图及其矩阵

由图 11.4 可见,3 和 4 这两个事件共同拥有 2 个人,2 和 3 共同拥有 1 个人,其他情况也可以从图中分析出来。

问题在于:上述两种矩阵尽管联系紧密,但是不可轻易进行比较。因此,我们需要一种算法找到二者之间的关系。为了比较二者之间的关系,需要构造如表 11.4 所示的个人—群体邻接矩阵(person-to-group adjacency matrix),该矩阵一般是最初调查得到的数据。

表 11.4　个体—群体邻接矩阵

A=

	1	2	3	4	5
A	0	0	0	0	1
B	1	0	0	0	0
C	1	1	0	0	0
D	0	1	1	1	1
E	0	0	1	0	0
F	0	0	1	1	0

在该矩阵中,列代表群体,行代表个人,如果元素 A_{ij} 等于 1,即表示第 i 行所代表的个人隶属于 j 这个群体。此时,从矩阵中看两个人共同参与的群体有多少,只需要比较各个行即可:找出两行对应值都是 1 的点,进行汇总,即可以计算出重叠。例如,如果比较 A 和 F,可以看出:A 隶属于群体 5, F 隶属于两个群体 3 和 4,但是他们却没有共同参与一个群体,即在对应值上不存在 1,二者没有重叠,如表 11.5 所示。

表 11.5　A 和 F 共同参与的群体的重叠情况计算

	1	2	3	4	5		∑
A	0	0	0	0	1	=	1
F	0	0	1	1	0	=	2
A×F	0	0	0	0	0	=	0

如果比较 D 和 F(如表 11.6 所示),可以看出:D 隶属于群体 2, 3, 4, 5, F

隶属于两个群体 3 和 4，他们共同参与的群体有两个，即 3 和 4，如表 11.6 所示。

表 11.6　D和F共同参与的群体的重叠情况计算

	1	2	3	4	5		∑
D	0	1	1	1	1	=	4
F	0	0	1	1	0	=	2
D×F	0	0	1	1	0	=	2

对于群体的分析来说也是如此。例如，群体 1 有两个成员，群体 2 也有两个成员，它们之间重叠的成员数是 1，如表 11.7 所示。

表 11.7　群体 1 和群体 2 中成员的重叠情况计算

	1	2	1×2
A	0	0	0
B	1	0	0
C	1	1	1
D	0	1	0
E	0	0	0
F	0	0	0
∑	2	2	1

如何把上述思想进一步推广，得到具有推广意义的结论呢？一般来说，我们可以计算出任何一对群体(个人)之间的重叠量，通过汇总个人—群体邻接矩阵(persons-to-groups adjacency matrix)中对应的行(列)对应因子之积即可。也就是说：个人—个人(Person)矩阵 $P_{ij}=\sum_{k=1}^{g}A_{ik}A_{jk}$。群体—群体(Group)矩阵 $G_{ij}=\sum_{k=1}^{p}A_{ki}A_{kj}$。

利用矩阵算法，可以很容易地计算上述矩阵。首先，定义 A^{T} 为矩阵 A 的转置矩阵，即 $A_{ij}^{T}=A_{ji}$。如果 A 的规模为 P×G，那么 A^{T} 的规模就是 G×P。因此，$P=A^{T}A$；$G=AA^{T}$，如表 11.8 所示。

表 11.8　个体共享群体、群体共享个问题矩阵

$A_{6\times5}=$

	1	2	3	4	5
A	0	0	0	0	1
B	1	0	0	0	0
C	1	1	0	0	0
D	0	1	1	1	1
E	0	0	1	0	0
F	0	0	1	1	0

$A_{5\times6}^{T}=$

	A	B	C	D	E	F
1	0	1	1	0	0	0
2	0	0	1	1	0	0
3	0	0	0	1	1	1
4	0	0	0	1	0	1
5	1	0	0	1	0	0

$P=A\times A^T=$

	A	B	C	D	E	F
A	1	0	0	1	0	0
B	0	1	1	0	0	0
C	0	1	2	1	0	0
D	1	0	1	4	1	2
E	0	0	0	1	1	1
F	0	0	0	2	1	2

$G=A^T\times A=$

	1	2	3	4	5
1	2	1	0	0	0
2	1	2	1	1	1
3	0	1	3	2	1
4	0	1	2	2	1
5	0	1	1	1	2

从理论上讲，这两个等式就界定了布雷格所说的“二元性”：“就网络成员而言，成为一个图(即矩阵 P)中的行动者同样可被看成是在其对偶图(dual picture)

表 11.9　4 个分支学会的成员及分支学会共享成员矩阵

A=

会员	城市	方法	网络	理论
1	1	0	0	0
2	0	1	0	0
3	0	1	0	0
4	1	0	1	0
5	1	1	0	1
6	0	1	1	0
7	0	1	0	0
8	1	0	1	0
9	0	0	1	1
10	0	0	0	1
11	1	0	0	0
12	0	0	0	1
13	0	0	1	1
14	1	0	1	1
15	0	0	1	1
16	0	1	0	1
17	0	1	0	0
18	0	0	0	1
19	0	1	1	0
20	0	0	1	0
21	1	1	1	0
22	1	0	0	1
23	0	1	0	0
24	1	0	1	0
25	0	0	1	1
26	1	0	1	0
27	1	0	0	0
28	1	0	0	1
29	0	0	1	0
30	0	1	0	0
31	1	1	0	0
32	0	0	1	0
33	0	0	1	1
34	1	0	1	0
35	1	0	1	1
36	0	0	1	1
37	1	1	0	0
38	0	0	1	1
39	0	0	0	1
40	0	0	1	1
41	1	1	0	0

$G=(A^T)A$

	城市	方法	网络	理论
城市	17	5	8	5
方法	5	14	3	2
网络	8	3	21	10
理论	5	2	10	18

(即矩阵 G)中的关系,对于群体来说也是如此”(Breiger, 1974:87)。计算结果得到的网络具有如下特点:(1)总是对称的;(2)对角线将告诉我们一个人隶属于多少个群体,或一个群体拥有多少个成员。

在一些情况下,此类计算具有一定的现实意义。例如,当我们收集到有 41 位学者参与中国社会学会下面的四个二级分会(城市社会学、社会学方法、社会理论和社会网络)这一数据(如表 11.6 所示)之后,很可能希望进一步了解,每个分会有哪些成员(这容易处理,只要看每个分会所在列中有多少“1”即可),各个分会之间有多少人是共享的。这种想法就可以利用上述矩阵计算的思想得以实现。我们可以利用 UCINET 进行这些操作,用矩阵 A 的转置再乘以矩阵 A,即可计算出各个分会之间有多少人是共享的,结果如表 11.9 所示。

读者完全可以根据自己收集的网络数据进行类似的计算。

二、二部“2-模图”分析

分析 2-模网络的上述方法是矩阵方法,除此之外,表征网络数据的方法还有图形法。因此,我们可以利用图形表示 2-模网络数据,这就需要首先构建二部矩阵(bipartite matrix),在此基础上建构二部图(bipartite “two-mode” graphs)。在 UCINET 中,利用 Data→Display 就可以展示 2-模网络的 2-部数据结构(bi-partite data structures)。

(一) 二部矩阵的构建

要构建一个二部矩阵,需要在隶属关系矩阵的基础上再加入一些行和列,使之变成方阵。这种矩阵不同于上述的 1-模矩阵,因为二部矩阵同时表达了两个模之间的关系,而上述 1-模矩阵表达的只是一个模态(集合)之间的关系。具体地说,上述学生—小组数据对应的二部矩阵将拥有 21 行 21 列(14 个同学加上 7 个小组),而两个 1-模网络的规模分别是 14 和 7。二部矩阵中的两个行动者—事件块与初始矩阵完全一样,两个新的矩阵(同学—同学矩阵和小组—小组矩阵)通常编码为 0。在 UCINET 中,点击 Transform→Bipartite,选择 affiliation. 14.7 数据,就可以把这个 2-模矩阵转换为 2-模二部矩阵。其对话框如图 11.5 所示。

在该对话框中,“value to fill within-mode ties”(即指定在每个模内部关系值是多少)一项通常设定为 0。因此,行动者之间(或者学习小组)只有在共同参与一些“学习小组”(或者有共同的学生)的时候才存在联系。因此,我们一般设

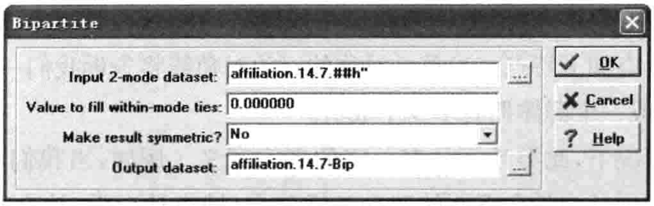


图 11.5　二部矩阵分析对话框

定每个模内部的关系是不存在的。点击 OK 后得到结果如表 11.10 所示。

表 11.10　二部矩阵分析结果

CONVERT 2-MODE TO BIPARTITE 1-MODE REPRESENTATION

Input dataset: "H:\Ucinet Data\DataFiles\affiliation.14.7

		1 A	2 B	3 C	4 D	5 E	6 F	7 G	8 H	9 I	10 J	11 K	12 L	13 M	14 N	15 m	16 p	17 e	18 g	19 c	20 s	21 b
1	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
2	B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
3	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1
4	D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
5	E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
6	F	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
7	G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
8	H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
9	I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
10	J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
11	K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0
12	L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
13	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1
14	N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0
15	maths	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	politics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	english	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	geography	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	chemistry	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	speech	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	biology	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

只要数据以这种方阵的形式表达,用来分析 1-模网络数据的方法都可用来分析该数据。但是在解释的时候需要注意,因为有待分析的网络是一类特殊的网络,其中的关系是在不同层次(学生和学习小组)上的点之间的关系。在一定意义上,行动者(学生)和事件(学习小组)可被看成是在单一分析层次上的社会对象(social objects),可以进而探讨其中心性和关联性等特征。但是,这种分析相对较少,尽管分析的结果可能出现某种有趣的结论。

通常情况下进行的操作是把行动者和事件"分开"处理,并且关注行动者是如何把事件关联在一起的,事件又是如何把行动者联系在一起的。下面我们将考察用来完成这种任务的技术,第一步是对数据进行可视化处理,即进行二部 2-模分析。

(二) 二部图的构建

在利用图来表达 2-模数据的时候不存在什么新的技术问题。行动者和事件

都被看成是点，线用来表达行动者与事件的关系（在行动者和行动者之间，或事件和事件之间不存在直接的关系）。

在 UCINET 中，点击 Visualize→NetDraw→File→Open→UCINET dataset→2-Mode Network 工具可生成二部 2-模图（Bipartite "Two-Mode" graphs）。图 11.6 展示了对上述 21 点图进行这种图形分析的结果。

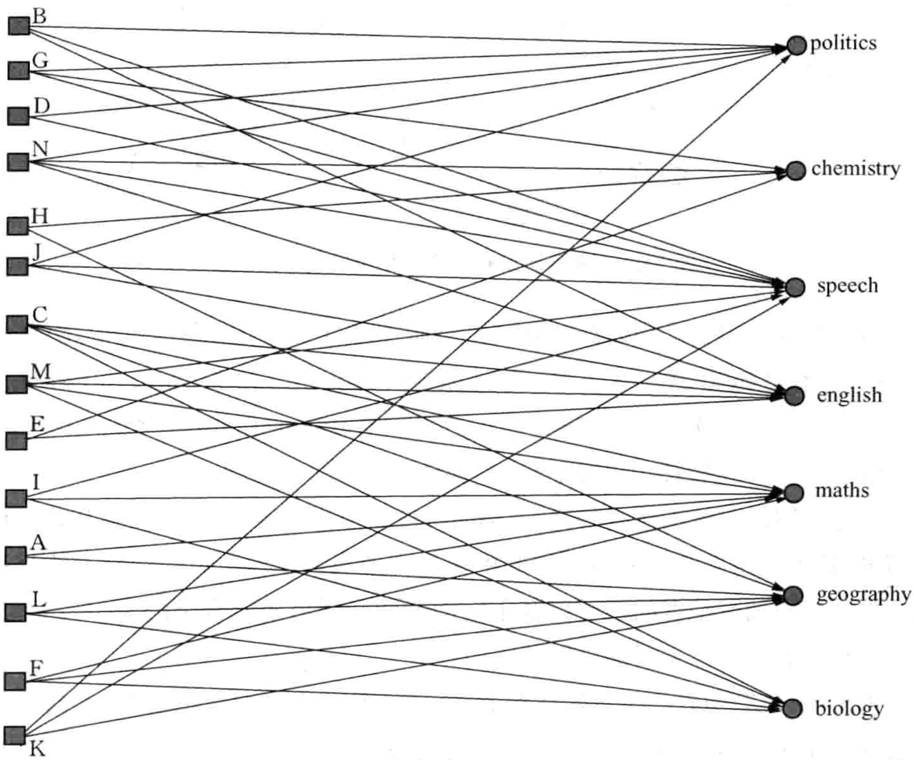


图 11.6　14 位同学参与 7 个小组的二部图分析结果

该图包含 21 个点，表现了一定的聚类性。对 2-模网络进行可视化处理的好处是可以清楚地看到 2-模关系的结构。除此之外还可能获得一些发现，特别是在利用某种量表方法（scaling method）来确定各个点的位置的时候更是如此，因为量表方法都要求距离近的行动者的关联度也高，因为它们拥有相似的事件截面（profiles of events）。例如，在画出上述图之后，点击 Layout→Graph－Theoretic layout→MDS（MultiDimensional Scaling，即多维量表）方法重新划图，可以得到如图 11.7 所示的结果。

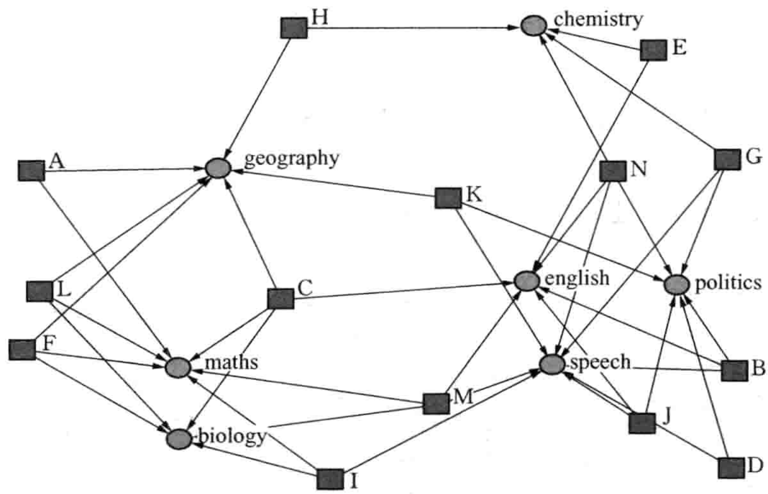


图 11.7　14 名同学参与 7 个小组的多维量表

由于多维量表技术是根据"距离"进行画图的,因此,在二维空间中,距离越近的点关系也越紧密。从上图可见,英语(english)、演讲(speech)和政治(politics)在距离上比较接近,这说明这三者具有较大的相似性。数学(maths)和生物(biology)比较接近,这说明这二者具有较大的相似性。

上述画图的方法固然有其优点,例如可以形象化地表示关系,但是其缺点也很明显,即不能给出关于关系结构的更多量化说明。下面介绍的计量方法将揭示由事件带来的行动者的聚类以及由行动者的共同出现所带来的事件聚类,从而揭示行动者—事件关系的深层次结构。

在 UCINET 中有一个自带数据 Davis。这是一个经典数据,即美国南方妇女数据(the Southern Women's data),该数据是戴维斯及其同事在 20 世纪 40 年代收集到的有关 18 位女性参与 14 个社会事件的数据。读者可就该数据进行上述 2-模网络分析。

第二节　2-模数据的定量分析

一、2-模数据的中心性分析

(一) 2-模数据的中心性分析

2-模网络的中心度分析比较复杂,下面主要结合 2-模隶属网进行说明。对

隶属矩阵的中心性分析的基本思路是，首先将隶属数据转换成“二部矩阵”，然后分析各种中心性指数。由于隶属网络中的行动者必须通过行动者所隶属的“事件”才能建立联系，因此，相关中心性研究要复杂得多。

1. 度数中心度。在二部图中，一个点的度数中心度是该点所隶属的事件数，一个事件的度数中心度是该事件所拥有的行动者数。这些量也等于行动者共享成员矩阵 X^N（行动者矩阵）中的对角线上的值（即隶属矩阵 A 的行总和），或者是事件叠加矩阵 X^M（事件矩阵）中的对角线上的值（或者隶属矩阵 A 的列总和）。

也可对这些中心度指标进行标准化处理。例如，一个点所隶属的事件数除以事件总数，就得到该点的相对度数中心度。一个事件的参与者人数除以参与人总数，就得到该事件的相对度数中心度。

2. 接近中心度。对于1-模网络中的一个行动者来说，该点到其他点的距离之和便是其接近中心度。而对于隶属网络中的一个点来说，这个距离之和等于该点与其他点的距离之和，再加上该点到所有事件的距离之和。然而，由于隶属网络是二分图网络，行动者仅仅与事件有关联，所有从一点发出的途径必然首先要经过该行动者所隶属的各个事件；同样，由于事件仅仅与行动者有关联，因此，所有从事件发出的途径也必然首先经过该事件所包含的各个行动者。

也就是说，在隶属网络中，一个行动者（点）n_i 的接近中心度是该行动者所隶属的事件到其他行动者和事件的距离的一个函数（Faust，1997：172—174），即

$$C_C^{NM}(n_i)=\left[1+\frac{\sum_{j=1}^{g+h}\min\limits_k d(k,\ j)}{g+h-1}\right]^{-1}$$

（其中 g 是点数，h 是事件数，事件 k 与点 i 邻接）

与之类似，一个事件的接近中心度也是该事件所含的行动者到其他行动者和事件的最短距离的一个函数，即

$$C_C^{NM}(m_k)=\left[1+\frac{\sum_{i=1}^{g+h}\min\limits_j d(i,\ j)}{g+h-1}\right]^{-1}\text{（其中行动者 } i \text{ 与事件 } k \text{ 邻接）}$$

从上述公式可以看到，在一个隶属网络中，行动者 n_i 的接近中心度和事件

m_k 的接近中心度之间存在联系。

3. 中间中心度。在1-模网络中,中间中心度关注的是一个行动者在多大程度上居于网络中间。由于在计算中间中心度的时候需要考虑到所有的捷径,因此,似乎不太可能把一个图中某点(行动者或事件)的中间中心度表达为该点邻接的其他点的中间中心度的一个函数。这里仅介绍隶属网络中二分图的行动者和事件的中间中心度。在隶属网络中,由于每对行动者之间的联络都要通过行动者参与各种事件来完成,因此,事件总是处于行动者之间的捷径上。与之类似,各对事件之间的联络都要通过行动者来完成,行动者也因而总处于事件之间的捷径上。

在一个隶属网络中,当计算一个事件 m_k 的中间中心度的时候,需要考虑所有属于该事件的行动者。事件 m_k 位于属于该事件的所有行动者的捷径上。如果给定的一对行动者(n_i, n_j)仅仅共享一个事件 m_k(即 $x_{ij}^N=1$),则 m_k 的中间度增加1个单位。如果给定的一对行动者(n_i, n_j)共享 x_{ij}^N 个成员,那么对于 m_k 中的任何一对行动者(n_i, n_j)来说,m_k 的中间中心度增加 $1/x_{ij}^N$ 个单位。这样看来,事件 m_k 的中间中心度比例可以表达为其每对成员共享资格的数目,即

$$\frac{1}{2}\sum_{ni,\, nj \in mk}\frac{1}{x_{ij}^N}$$

从这个公式中可以看到,对于一个事件来说,只有当该事件的每对成员在该事件中相遇时,该事件才获得中间中心度。

另外,如果一个行动者仅仅归属于一个事件,则该事件也获得中间中心度。在这种情况下,所有来自这个行动者的捷径都包含该事件。在一个隶属网络中,由于存在 $g+h$ 个点(行动者和事件),对于不属于其他事件而只属于一个事件的每一个行动者来说,该事件得到 $g+h-2$ 个中间"点"。

所有这些观测值说明,只有当一个事件的行动者不属于其他事件的时候,该事件的中间中心度才增加。由于拥有这种单成员资格的行动者(single-membership actors)不处于任何事件对(pairs of events)之间,因此,一个事件一般在包含"非核心行动者"(non-central actors)的时候,该事件的中间中心度才增加。同理,当多对行动者仅仅共享一个事件的时候,该事件的中间中心度也增加。

行动者的中间中心度的性质也类似。如果一个行动者仅仅是一个事件的唯一成员,则该行动者得到 $g+h-2$ 个中间"点"。对于一个行动者所隶属的所有

"事件对"来说,该行动者将获得 $1/x_{kl}^{M}$ 个中间中心度(Faust, 1997:175)。

下面,我们以 UCINET 自带的 2-模数据 Galask 为例,展示 2-模数据的中心性分析过程。Galask 数据是由管理学家 J.加拉斯科维茨(J.Galaskiewicz)在 1978—1981 年间收集的 26 行 15 列的隶属关系数据,它记载了 26 名 CEO 参与 15 家俱乐部的情况。

在 UCINET 中,点击 Network→Centrality→2-Mode Centrality,在"Input 2-Mode matrix"(即需要指定输入的 2-模矩阵)中选择数据 Galask,点击 OK 后得到如表 11.11 所示的结果。

表 11.11 Galask 数据的 2-模中心性分析结果

2-MODE CENTRALITY

2-Mode Centrality Measures for ROWS

(为求简洁,第 2, 5, 6, 9, 13, 15—19, 22, 23, 25 行略)

	Degree	Closeness	Betweenne
1	0.200	0.730	0.013
3	0.133	0.670	0.005
4	0.200	0.684	0.004
7	0.267	0.747	0.049
8	0.200	0.556	0.012
10	0.133	0.670	0.004
11	0.200	0.684	0.010
12	0.133	0.546	0.004
14	0.467	0.802	0.067
20	0.333	0.765	0.032
21	0.200	0.730	0.019
24	0.333	0.765	0.051
26	0.200	0.684	0.008

2-Mode Centrality Measures for COLUMNS

	Degree	Closeness	Betweenne
1	0.115	0.482	0.007
2	0.423	0.600	0.069
3	0.846	0.871	0.488
4	0.462	0.643	0.182
5	0.115	0.482	0.005
6	0.154	0.491	0.008
7	0.154	0.474	0.016
8	0.154	0.491	0.017
9	0.231	0.519	0.022
10	0.115	0.466	0.007
11	0.154	0.491	0.007
12	0.192	0.500	0.014
13	0.192	0.509	0.016
14	0.115	0.443	0.003
15	0.346	0.574	0.049

左表给出了 26 位 CEO 的中心度分析结果。从中可见,度数中心度最高者是第 14 个 CEO,其值为 0.467;度数中心度最低者是第 3、第 10 和第 12 个 CEO,他们的度数中心度值都是 0.133。也可以看出,度数中心度最高者,即第 14 个 CEO 的接近中心度和中间中心度也最高,这说明他毫无争议地成为网络

的核心人物。但是,在度数中心度最低的第3、第10和第12个CEO中,只有第12个CEO的接近中心度和中间中心度也最低,这说明第12个CEO是该隶属网络的边缘人物。

右表给出的是15个俱乐部的中心度分析结果。可见,第3号俱乐部的度数中心度、接近中心度和中间中心度都最高,其度数中心度值达到0.846,接近中心度是0.871,中间中心度是0.488。这说明该俱乐部处于隶属网络事件的核心。第14号俱乐部的度数中心度、接近中心度和中间中心度最低,这说明该俱乐部位于隶属网络事件的边缘。

上述分析程序给出了Galask数据的三种中心度指标,这些都是标准化了的指标,便于比较。除此之外还有特征向量中心度指标、流量中心度指标(flow betweeness centrality)等,UCINET程序没有计算这些指标。感兴趣的读者可进一步参见福斯特(Faust, 1997)。

(二) 2-模数据的中心度可视化分析

利用各种中心度指数可以揭示每个点的中心程度。除此之外,我们还可以对中心度进行可视化分析,使得中心度大的点的规模也大。这样的话,每个点到底在多大程度上居于网络的核心,在图中就一目了然。这种想法可以在UCINET的NetDraw中得以实现。在NetDraw中打开"学生—学习小组"数据之后,会出现如图11.8所示的数据图。

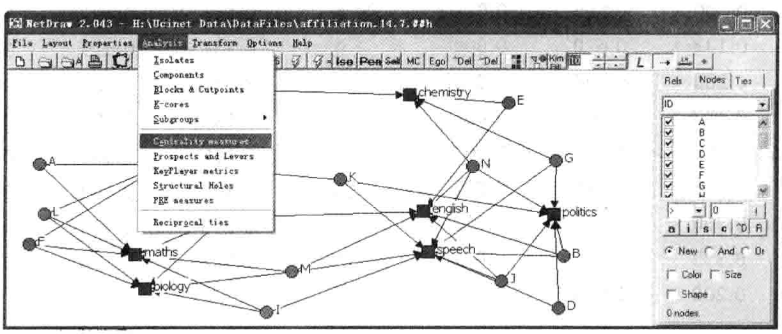


图 11.8 中心性测度对话框

然后点击Analysis→Centrality Measures后会出现对话框,其中可以选择的中心度指数有六项(即Measures下面的六项),在"Set Node Sizes By"中选择中间中心度,结果如图11.9所示。

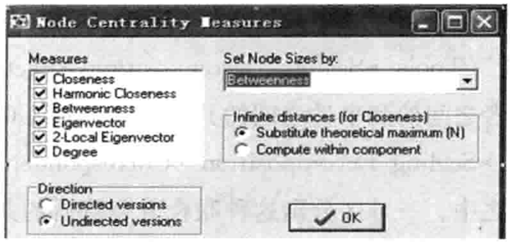


图 11.9　点的中心度测度对话框

从图 11.10 中可以清楚地看到哪些点居于网络的核心。例如,“speech”(演讲)点最大,说明就“中间中心度”这个指标而言,它居于网络的核心。

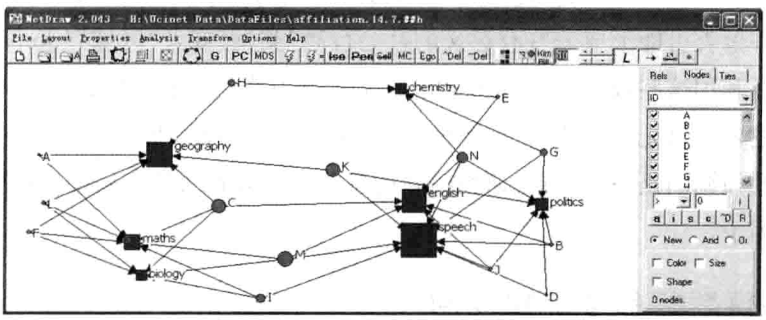


图 11.10　根据中心度重新画的图

当研究多个变量的时候,我们往往希望找到它们之间共享的特征。此时,可利用有关“协变”(co-variation)方面的信息,从而推导出这些变量背后的一些维度和因子。关注该维度和因子就是因子分析和成分分析的任务,这两种分析方法关注变量之间共同的因子以及每个个案在各个维度上的取值大小。

我们可以利用同样的“量表”(scaling)逻辑来分析有关行动者与事件之间的关系网络,即可以对 2-模网络进行因子分析。可以根据行动者参与事件的情况来表征它们之间的相似性,但是要根据它们之间的共同方差对事件进行加权分析。与之类似,也可以对事件进行指标化处理,这要根据行动者共同参与的模式来分析——但是要根据它们之间共同参与的频次对行动者进行加权。诸如多维量表、因子分析和主成分分析等都可用来对事件或行动者进行计量。

在利用这种计量思路分析行动者—事件之间关系数据的时候,可采用 UCINET

中的一些技术,包括“奇异值分解”(Tools→Scaling/Decomposition→SVD)和“2-模量表因子分析”(Tools→Scaling/Decomposition→Factor Analysis),它们都将同时考察行动者之间的和事件之间的共享方差。在UCINET中还包括对应分析技术(Tools→Scaling/Decomposition→Correspondence),它将同样的逻辑应用到二值数据之中。一旦区分出这种联合方差背后的共同维度,就可以把行动者和事件“画在”同一个空间之中,进而看出哪些行动者在参与事件这个问题上是相似的,哪些事件在有行动者参与这个问题上是相似的,哪些行动者和事件在位置上相互接近。

在某些情况下还可以通过解释背后的因子或维度获得一些洞见,能够深刻地认识到为什么行动者和事件以一定的方式相互接近。一般来讲,行动者和事件的某些聚类可能形成一些有意义的社会行动“类型”或者“区域”(domains)。

下面利用这些技术简要分析上述学生—学习小组数据,以便说明2-模数据分析的逻辑。首先介绍奇异值分解,然后介绍因子分析,最后探讨对应分析技术。需要强调的是,这些方法研究的目的都是为了降低维度,从而找到数据背后的共同因子。

二、2-模数据的奇异值分解

奇异值分解法(Singular Value Decomposition, SVD)是一种用来找到2-模(多值)网络数据背后因子的方法。为了展示SVD,我们需要输入一个2-模矩阵。

SVD的原理如下:对于一个n×m矩阵(n不小于m)X来说,SVD算法将找到矩阵U、D和V,使得X= UDV′(V′是V的转置矩阵)。矩阵D是一个r×r的对角线矩阵,它包含r个奇异值。矩阵U是一个n×r矩阵,它包含矩阵XX′的r个特征向量(eigenvectors); V是一个m×r矩阵,它包含矩阵X′X的r个特征向量。各个特征值按照降序排列。对于对称数据来说,U和V相等。

用矩阵的语言表示SVD定理就是:秩[①]为r的实矩阵A必有如下分解式:

---

① 矩阵的秩是矩阵的列向量组(或行向量组)的任一极大线性无关组所含向量的个数。通俗地讲,矩阵的秩指的是矩阵中相互独立的向量数的最大值,它表征了矩阵中独立向量的维度,相当于数据矩阵的背后共同的维度。

$$A = U\begin{pmatrix} \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \cdots & \\ & & & a_r \end{pmatrix} & 0 \\ 0 & 0 \end{pmatrix} V'$$

该公式就叫做 A 的奇异值分解。其中，U 和 V' 都是正交阵①，并且 $a_i > 0$；$i = 1, 2, \cdots, r$。$a_1, a_2, \ldots, a_r$ 叫做 A 的奇异值。

由于

$$A'A = V\begin{pmatrix} \begin{pmatrix} a_1^2 & & & \\ & a_2^2 & & \\ & & \cdots & \\ & & & a_r^2 \end{pmatrix} & 0 \\ 0 & 0 \end{pmatrix} V';\ AA' = U\begin{pmatrix} \begin{pmatrix} a_1^2 & & & \\ & a_2^2 & & \\ & & \cdots & \\ & & & a_r^2 \end{pmatrix} & 0 \\ 0 & 0 \end{pmatrix} U'$$

所以，奇异值就是 $A'A$（也是 AA'）的非零特征根的平方根。

进行 SVD 分析的目的是什么？可以解释如下。当我们面对的隶属关系数据的行数和列数较多的时候，可通过奇异值分解的技术，降低其维度数，以便找到数据背后共同的因子，这就是 SVD 的目的。该方法将分解出一些奇异值(singular values)，这是一种不同于常规的因子分析和主成分分析的方法。因此，在分析 2-模网的结构时，一个比较好的想法就是应该既考虑到 SVD，又考虑 2-模因子分析的结果。

在 UCINET 中，点击 Tools→Scaling/Decomposition→SVD 后，出现对话框。需要注意的是，输入数据的行数不能少于列数（一般要求行数要远远大于列数，否则需要先对数据矩阵进行转置，再进行运算），并且矩阵最好是多值矩阵（对于二值矩阵来说，最好应用下面将介绍的对应分析法来分析）。在出现的对话框中有一些选项，其中有三个选项要求对行和列的值进行计量：

1. 坐标(Coordinates)：每个特征向量根据与之对应的特征值来加权处理。

2. 负载值(Loadings)：每个特征向量根据与之对应的特征值的平方根来加

① 正交阵的定义如下：设 A 是 n 阶实阵（其中的值都是实数），并且 $AA' = I_n$，则称 A 为正交阵。这里的 I_n 为 n 阶单位阵。

权处理(当利用 SVD 来分析相关系数矩阵的时候,这个选项产生的是因子负载值)。

3. 轴(Axes):不进行重新计量,这是默认值。

需要保存的因子的个数(No of factors to save):默认值为 3 个。研究者可根据情况进行修改,但是其最大值为 r,即用来分解矩阵 X 的特征向量的数目。

是否根据各个因子对矩阵进行重构(Reconstruct matrix from factors):默认值为“否”。如果选择“是”(YES),那么程序会根据 r 个特征向量计算出 UDV′这个乘积。其结果是根据“最小二乘法”标准(least squares criterion)利用秩为 r 的矩阵对 X 的最佳逼近。

输出的数据有多项。其中一项为包含行值的文件,名为 RScores。它是包含 U 矩阵的文件。另外一项输出文件为包含列值的文件,名为 CScores,它是包含 V 矩阵的文件。第三个输出的文件为包含奇异值的文件,名为 Eigen,它是包含 D 矩阵的文件。第四个输出的文件为包含重构的矩阵的文件,名为 Recon。它是包含与 X 逼近的矩阵 UDV′的文件。最后一个输出的文件为行和列联合在一起的矩阵文件,名为 RCScores,它将 U 矩阵和 V 矩阵连接在一起,产生一个 (M+n)×r 的矩阵,可把它的行和列画在一个图中。

如果在上图中的“Input Data”一项中选择我们给出的学生—小组数据,那么经过计算,上述学生、小组矩阵的秩(rank)为 7,因而出现 7 个奇异值。对该矩阵进行 SVD 分析的结果如表 11.12 所示。

表 11.12 奇异值

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO	PRE	CUM PRE
1:	4.379	30.8	30.8	1.289	0.085	0.085
2:	3.399	23.9	54.7	1.677	0.484	0.569
3:	2.027	14.3	69.0	1.061	0.172	0.741
4:	1.911	13.4	82.4	1.403	0.153	0.894
5:	1.361	9.6	92.0	2.102	0.078	0.972
6:	0.648	4.6	96.5	1.314	0.018	0.990
7:	0.493	3.5	100.0	1.314	0.010	1.000
	14.217	100.0				

在分析得到的 7 个奇异值中,第一个奇异值是 4.379,占总体的 30.8%,第二个奇异值是 3.399,占总体的 23.9%,累计百分比为 54.7%。两个奇异值是比较重要的。这些奇异值的含义是什么?这一点可以通过针对小组的进一步 SVD 数据分析的结果来说明,如表 11.13 所示。

表 11.13　对学生—小组数据的 SVD 分析的结果(小组量表)

Column Scores

		1	2	3
1	maths	0.362	-0.520	-0.164
2	politics	0.397	0.440	0.163
3	english	0.415	0.135	-0.244
4	geography	0.303	-0.412	0.671
5	chemistry	0.199	0.189	0.552
6	speech	0.548	0.345	-0.225
7	biology	0.328	-0.439	-0.287

可以认为,第一个维度(奇异值)反映的是这些同学的共同能力,第二个维度反映了学生在“文科”(政治、英语)方面的能力(因为该奇异值在政治、英语、演讲等方面的负载值大);第三个反映了学生在“理科”(地理、化学)方面的能力。因此,我们可以命名第二个奇异值为“文科因子”,第三个奇异值为“理科因子”。可见,演讲小组、政治小组、英语小组居于一个维度,化学小组、生物小组、地理小组等处于另外一个维度。

结合这些奇异值,我们就可以考察行动者和事件在每个奇异值上的负载值是多少。首先,从计算结果中可以看到每个行动者在各个维度上的负载值,分析结果如表 11.14 所示。

表 11.14　对学生—小组数据的 SVD 分析结果(学生量表)

Row Scores

		1	2	3
1	A	0.152	-0.274	0.250
2	B	0.310	0.271	-0.151
3	C	0.321	-0.364	-0.012
4	D	0.216	0.231	-0.031
5	E	0.140	0.095	0.152
6	F	0.227	-0.404	0.109
7	G	0.261	0.287	0.242
8	H	0.115	-0.066	0.603
9	I	0.283	-0.181	-0.333
10	J	0.310	0.271	-0.151
11	K	0.285	0.110	0.300
12	L	0.227	-0.404	0.109
13	M	0.377	-0.141	-0.454
14	N	0.356	0.326	0.121

从结果中可以看出,A 同学在“文科因子”上的负载值为−0.274,而在“理科因子”上的负载值为 0.250。这说明,A 擅长于理科,不擅长文科。

在常见的因子分析和主成分分析中,“奇异值”(singular values)与“特征值”(eigenvalues)类似。上述计算也表明,奇异值是非零特征根的平方根。这里的计算结果说明,有关行动者和事件之间的联合方差“空间”(joint space of the variance)不能仅仅根据一个奇异值来描述,还需要至少另外一个维度来解释。当

然,如果我们清楚地认识到数据背后存在着诸如"左派/右派"、"金钱/面子"等简单分组的话,那么仅利用少数奇异值就足以解释联合方差的大部分。在本案例中,结果告诉我们,用来解释行动者和事件的联合方差的维度至少不是单一的。

有时候需要对各个行动者和事件的位置进行可视化处理,需要用到散点图(scatterplot),在 SVD 计算得到的输出文件中会给出二维散点图,它可以记载前两个特征向量之间的关系(当然也可以描述第一个和第三个特征值之间的关系或者第二个和第三个特征值之间的关系,但这需要研究者自己在输出的图中进行选择)。可以对该散点图进行编辑、保存、整理、打印等操作,例如可以加入或者省略标签。保存之后的图像可利用 Tools→Scatterplot 来查看,也可利用 Word 处理程序读取。

就本例而言,如果根据前两个维度进行展示,可以得到如图 11.11 所示的散点图。

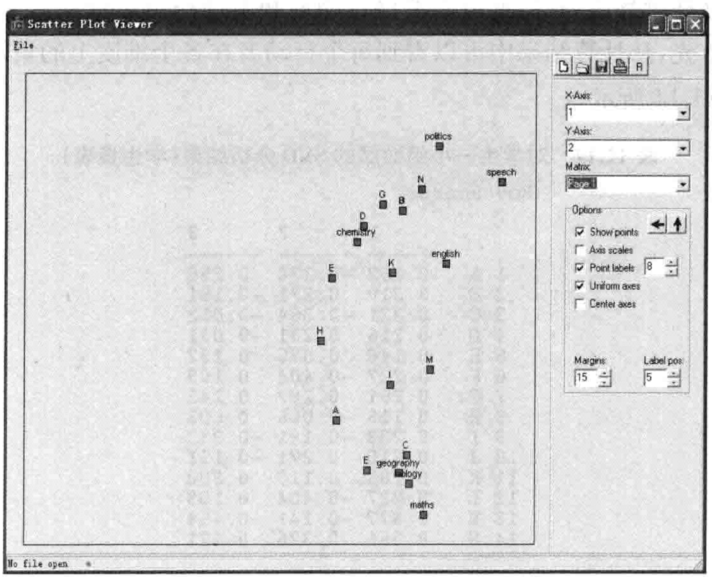


图 11.11　对学生—小组数据的 SVD 分析二维图

可见,从第一个维度(即图中的横轴,相当于同学们的"共同素质"维度)中看不出各个事件之间的明显差异,而第二个维度(即图的纵轴)差异明显。散点图并没有清楚地把特定的一些行动者和事件沿着强线性维度进行定位,但是它确实产生了一些有趣的聚类,表明行动者所在的群体,也表明了一些有关参与模式

的核心问题。例如,“geography”、“biology”和“maths”很靠近,明显聚在一起;而右上部分的“politics”、“speech”和“english”也相对比较靠近。

需要注意的是,针对 SVD 计算的结果不一定有明确的解释。另外,即使针对同一个特征根,接受不同学术训练的人给出的解释也可能不同。与此同时,我们也不能保证做出的解释完全符合数据。例如,在学生—小组数据的 SVD 分析结果(学生量表)中,M 同学在两个因子上的负载值都是负数,我们能据此简单地说 M 不具有文科因子和理科因子吗?实际上,由于他同时参与了数学、英语、演讲和生物这四个小组的学习,因此,我们也可以说 M 同时具有文科素养和理科素养。

三、2-模数据的因子分析

一谈到因子分析,往往与主成分分析关联在一起,二者异同点并存。在有些文献中,主成分分析法(principal components analysis, PCA)也被称为“因子分析法”,但实际上二者是不同的。一般认为,因子分析的含义比较广泛,用来进行因子分析的方法比较多,其中之一就是主成分分析。在 SPSS 中,主成分分析也是因子分析程序的一个部分,这也使之成为得到广泛应用的一种数据分析方法(郭志刚,1999; Scott, 2013:154—157)。

主成分分析是作为一种分析个案—变量(a case-by-variable)属性矩阵的方法发展起来的,其目的是为了发现多个变量中共享的一个或多个因子或成分。它试图利用原始数据发现一系列坐标或轴(因子或维度),用来绘制出数据散布图。根据个案—隶属项(a case-by-affiliation)关系数据得到的散布图具有如下性质,即从一个点到另一个点的空间距离和环绕方向表达了它们之间的实际相对位置。

一个简单的 PCA 算法将首先把个案—变量矩阵转换为变量—变量(variable-by-variable)矩阵,从而表明变量(即初始阵的各列)之间的相关关系。因此,新的矩阵展示了各个变量之间的相关性程度。下一步是搜索这个矩阵,找到代表了最高度相关的一组变量,并且用一个构造出来的人工变量代替,以此来测量各个变量之间的相关关系。如此看来,相关程度高于某个特定值的一组变量将用一个构造出来的变量来替代,这个构造出来的变量就称为第一主成分。接下来寻找另外一系列高度相关,但是与第一系列变量不相关的变量。用来代替这些变量的构造出来的新变量就是第二主成分。主成分分析法将持续此类过

程,一直到确定出所有可能的成分,这些主成分共同解释了数据中发现的全部变差。通过这种分析过程,初始的变量—变量相关系数矩阵就转换成为一个变量—成分相关矩阵。在主成分分析中,一般需要搜索出能够解释数据中较高比例方差的最少数目的主成分。

因此,PCA 开始于一个变量—变量相关系数矩阵,它是根据初始的个案—变量矩阵构造出来的。根据变量—变量矩阵可以构造出一个变量—成分矩阵,该矩阵的每个元素表达的是计算出来的每个变量在每个成分上的"负载值"(loadings)。把各个主成分当成一个散点图的各个轴,各个负载值用来确定每个变量在各个轴上的位置。

有时候需要对数据的结构进行旋转,目的是为了更清晰地揭示其结构。旋转的方法主要有两种:直角旋转法和斜角旋转法。其中直角旋转法又包括方差最大法(varimax solution)和四次方最大法(quartimax solution)。方差最大法是指在旋转的时候,让因子矩阵中每一列的值尽可能变为 1 或者 0,以便突出每一个因子的性质。该方法从简化因子矩阵的每一列出发,使得和每个因子有关的负载值平方的方差最大。当只有少数的几个变量在某个因子上有较高的负载值时,对因子的解释是最简单的。当与某个因子有关的负载值的平方的方差最大时,因子具有最大的可解释性(郭志刚,1999:104—105)。与之对应,四次方最大法的含义是,在旋转的时候令每一个变项在某个因子上的负载值尽可能为 1,而在其他因子上的负载值尽可能为 0,以便突出一个最强因子的性质。从计算的角度讲,该方法是从简化因子矩阵的行出发,通过旋转初始因子,使得每一个变量只在一个因子上有较高的负载值,在其他因子上有尽可能低的负载值。该方法通过使因子负载矩阵中每一行因子负载平方的方差达到最大求得因子解(郭志刚,1999:104)。

一般情况下,如果各个点在空间中沿着一个特定方向散布的话,那么转动这个结构,直到各个点沿着第一主成分最大限度地展开。旋转程序的目的是为了形成一个构型的定位,以便更好地与主轴吻合。旋转程序输出的结果是一个新的变量—成分矩阵,它包含了每个变量的一组修改过的负载值。

PCA 的分析结果给出了由一组维度界定的一个空间上多个变量的分布。按照类似的方法也可形成个案的散布图。在这种分析中,首先把初始的个案—变量矩阵进行转置,计算出一个个案—个案相关系数矩阵,最终计算出一个个

案—成分负载值矩阵，从而可以绘制出个案分布图来。一般将这种个案—个案分析称为“Q-模式”的主成分分析法，以便与常规性的变量—变量“R-模式”分析区分开来。

在用因子分析法处理关系数据的时候也有两种方式，即Q-模式因子分析和R-模式因子分析。通常情况下，PCA为R-模式的，即它针对初始阵的各列进行操作。因此，对一个个案—隶属项发生阵(case-by-affiliation incidence matrix)的R-模式的因子分析就是对数据矩阵各个隶属项的共同因子进行分析。如果研究者希望考察每个个案之间的关系结构，就有必要对这个矩阵进行转置，从而把个案所在的各行变为矩阵的各列，如在一种Q-模式的PCA分析中一样。例如，可以对一个个人—组织发生阵直接进行分析，从而形成一个组织—成分矩阵，或者把这个矩阵进行转置分析，从而形成一个个体—成分矩阵。

在分析“学生—小组”矩阵的过程中，PCA算法将考察各个小组，以便找到那些在其成员属性上最相似的学习小组。与CONCOR一样，此时也用相关系数作为相似性测度，所发现的各个具有相似性的小组集合将用一些主成分来代替。对初始发生阵实行转置就可以分析各个个案，从而进行Q-模式的因子分析。如前文几章所讨论过的CONCOR和块模型程序中所看到的那样，从同样的输入数据中产生的列解(column solution)和行解(row solution)将相互具有“对偶性”(dual)。它们虽然不同，但表征了同一组数据中的互补性(Scott, 2000:156)。

对于无向关系的对称数据矩阵来说，行解和列解将完全相同。但是对于一个有向邻接矩阵而言，两个解会有所不同，一个解可能对应于由“发送”关系生成的网络，另一个解则对应于由“接受”关系生成的网络。

(一) R-模式的因子分析(变量—变量分析，即针对隶属项之间的关系进行的分析)

在UCINET中，点击Tools→Scaling/Decomposition→Factor Analysis，即可执行因子分析。点击后出现对话框，选择上述2-模学生—小组数据进行因子分析，结果如表11.15所示。

一般认为，选取的因子数目应该使得累积解释的方法比例达到70%—80%以上。从计算结果中可以看到，前两个因子可以解释掉数据中总方差的70.5%；前三个因子可以解释掉数据中总方差的86.3%。因此，把因子数目确定为3是比较稳妥的。当然，也可以把因子数目确定为2，这样解释起来可能更容易一

表 11.15　针对学生—小组数据进行 R 模式的 2-模因子分析得到的特征值

```
FACTOR ANALYSIS
------------------------------------------------------------------------
Method of extraction:           PRINCIPAL COMPONENTS
Method of rotation:             VARIMAX
Minimum eigenvalue to retain:   1
Data type:                      2-mode raw data matrix
Input dataset:                  "H:\Ucinet Data\DataFiles\affiliation.14.7

EIGENVALUES

 FACTOR  VALUE PERCENT   CUM %   RATIO
 ------- ------ ------- ------- -------
      1: 3.447    49.2    49.2   2.314
      2: 1.489    21.3    70.5   1.347
      3: 1.106    15.8    86.3   1.913
      4: 0.578     8.3    94.6   2.668
      5: 0.217     3.1    97.7   2.338
      6: 0.093     1.3    99.0   1.310
      7: 0.071     1.0   100.0
 ======= ====== ======= ======= =======
         7.000   100.0
```

些。问题是,每个因子的含义是什么？这一点需要通过观察各个学习小组在因子上的负载值来解释,如表 11.16 所示。

表 11.16　旋转的因子负载值和每位同学在 3 个因子上的负载值

```
FACTOR ANALYSIS
---------------------------------------------------
Rotated Factor Loadings
                           1      2      3
                       ------ ------ ------
  1       maths        -0.831 -0.483 -0.082
  2    politics         0.390  0.867 -0.005
  3     english         0.040  0.033  0.898
  4   geography        -0.152 -0.570 -0.676
  5   chemistry         0.842 -0.244  0.232
  6      speech        -0.026  0.925  0.233
  7     biology        -0.838 -0.385  0.093

Factor scores
                1      2      3
             ------ ------ ------
  1 A        -0.134 -0.764 -1.192
  2 B         0.059  1.149  0.718
  3 C        -0.941 -1.200  0.621
  4 D         0.100  1.454 -0.726
  5 E         1.448 -1.323  1.462
  6 F        -0.900 -0.895 -0.823
  7 G         1.157  0.711 -0.290
  8 H         1.605 -1.335 -0.862
  9 I        -1.421  0.357  0.103
 10 J         0.059  1.149  0.718
 11 K         0.215  1.137 -1.607
 12 L        -0.900 -0.895 -0.823
 13 M        -1.463  0.052  1.547
 14 N         1.116  0.405  1.154
```

对结果的解释:可以将因子 1 解释为“文科因子”,因子 2 解释为“理科因子”,因子 3 解释为“英语因子”(显然,这样的解释未必恰当,这里暂且这样界定因子的含义)。每个学生在各个因子上的负载值体现了此人的能力。例如,D 同学的理科能力很强(负载值为 1.454),文科能力很弱(负载值为 0.1)。对其他学生的解释与之类似。

通过这种分析，就可以揭示关系数据背后的潜在因子，也对各位同学的文理科能力有了比较清晰的认识。

（二）Q-模式的因子分析（个案—个案分析，即针对行动者之间的关系进行的分析）

如果分析个案—个案之间是否存在共同因子，这种分析就是 Q-模式的分析，它是对 2-模矩阵的转置矩阵（转置矩阵可通过 UCINET 中的 transpose 命令执行）进行的分析。在 UCINET 中，首先对数据 affiliation.14.7 进行转置处理，得到的数据可命名为 affiliation14。然后，沿着 Tools→Scaling/Decomposition→Factor Analysis 路径，选择转置之后的"小组—学生"数据 affiliation14，点击 OK 后会计算出如表 11.17 所示的结果。

表 11.17　因子分析结果

```
FACTOR ANALYSIS
-------------------------------------------------------------
Method of extraction:      PRINCIPAL COMPONENTS
Method of rotation:        VARIMAX
Data type:                 2-mode raw data matrix
Input dataset:             "H:\Ucinet Data\DataFiles\affiliation14

     EIGENVALUES
        FACTOR      VALUE    PERCENT      CUM %      RATIO
      --------   --------   --------   --------   --------
            1:      6.988       49.9       49.9      2.348
            2:      2.976       21.3       71.2      1.260
            3:      2.362       16.9       88.0      2.072
            4:      1.140        8.1       96.2      3.438
            5:      0.332        2.4       98.6      1.639
           …  …  …  ……  …  …  …  …  …  …  …
      =======    =======    =======    =======    =======
                   14.000      100.0

     Rotated Factor Loadings
                            1          2          3          4
                     --------   --------   --------   --------
      1    A           -0.661     -0.196      0.322     -0.428
      2    B            0.275      0.132      0.170      0.937
      3    C           -0.945      0.172     -0.220     -0.151
      4    D            0.571      0.151      0.651      0.474
      5    E            0.194     -0.345     -0.876      0.245
```

6	F	- 0.698	0.177	0.205	- 0.656
7	G	0.945	- 0.172	0.220	0.151
8	H	0.057	- 0.821	- 0.108	- 0.473
9	I	- 0.008	0.897	0.201	- 0.345
10	J	0.275	0.132	0.170	0.937
11	K	0.151	- 0.302	0.871	0.281
12	L	- 0.698	0.177	0.205	- 0.656
13	M	- 0.254	0.892	- 0.224	0.160
14	N	0.698	- 0.177	- 0.205	0.656

Factor scores

		1	2	3	4
1	maths	- 0.806	0.904	0.060	- 0.831
2	politics	0.699	- 0.493	1.116	0.944
3	english	- 0.853	- 0.019	- 1.471	1.747
4	geography	- 1.284	- 1.524	0.958	- 0.524
5	chemistry	1.465	- 1.072	- 1.298	- 0.973
6	speech	1.108	0.971	0.943	0.554
7	biology	- 0.328	1.234	- 0.307	- 0.916

第一个因子的解释力最大,约占 50%;当因子达到 3 的时候,三个因子就可以解释掉数据中总方差的 88%。因此,我们可以将通过 Q-模式因子分析得到的因子数确定为 3。

问题是,这三个因子不同于上述 R-模式因子分析得到的三个因子,它们的含义是什么?这一点有时候没有明确的意义,因而难以解释。如果真需要解释的话,也需要结合各个学生的具体信息。例如,如果关注每个学生所在的班级,假设发现在因子 1 上负载值较高的学生多数来自一班,在因子 2 上负载值较高的学生主要来自二班,在因子 3 上负载值较高的学生多数来自三班,那么可以把这三个因子分别命名为“一班因子”、“二班因子”和“三班因子”。当然,实际数据中很少出现这种明显的分类情况。

这样看来,进行 Q-模式的因子分析的结果不一定有实际意义,本例就是如此。本例只是为了展示如何进行 Q-模式的因子分析,计算的结果很难解释,这一点请读者注意。如果研究者认为行动者(个案)之间存在着共同因子(例如分为各个班级,并且每个班级的同学参加学习小组的情况明显不同),那么可利用这种 Q-模式分析的结果来找到共同因子。

因子分析与 SVD 的目标是一样的，都是用来区分出行动者—事件方差(space of actor-by-event variance)的联合空间背后的维度，并且在该空间中对行动者和事件进行定位。用来区分不同维度的因子分析不同于 SVD。

因子分析的解不同于 SVD，它表明在学生和小组的联合方差方面存在复杂的维度。也就是说，对各个维度的简单描述并不足以提供非常精确的预测，不能预测每个行动者和事件的定位。尽管如此，因子分析法却提供了比 SVD 分析稍微简洁的结果，尽管该结果类似于 SVD 的结果，但是在具体细节上仍然存在重要的差别。

总之，我们已经有一定的印象，即三个因子的拟合度很低。为了对这种模式进行可视化分析，需要把行动者和事件在各个维度上的负载值从输出文件中抽取出来，利用 scatterplot 进行画图。

四、2-模网络的对应分析

对应分析最初是于 1970 年提出来的一种多变量统计分析方法，它把 R-型因子分析和 Q-型因子分析统一起来，通过 R-型因子分析直接得到 Q-型因子分析的结果，把变量(指标、事件、隶属项等)和行动者同时反映到相同的坐标轴(因子轴)的一张图上，用此来说明隶属项与行动者之间的对应关系。

对于二值数据来说，一般不推荐利用 SVD 和因子分析。这是因为，因子分析的基础是行动者和事件之间的相关系数矩阵或者协方差/方差矩阵。当用二值数据表示行动者与事件之间关系的时候，相关系数可能不能充分地表示协方差，从而使得关系模式难以显现。因此，严格地讲，上文利用 SVD 和因子分析来分析二值学生—小组数据是不太恰当的。

(一) 对应分析的原理

对应分析(correspondence analysis)是在因子分析的基础上发展起来的。上文已经指出，因子分析分为 R-型因子分析和 Q-型因子分析。R-型因子分析是对变量(指标、事件、或者隶属项等)作因子分析，研究的是变量(指标、事件、或者隶属项等)之间的相互关系，即各列之间的关系；Q 型因子分析是对行动者(样品、个案等)作因子分析，研究的是各个行动者即各行之间的相互关系。在因子分析的结果中可以看出变量之间的接近程度。然而，在错综复杂的经济和管理关系中，人们所关心的不仅仅是行或列本身变量之间的关系，还关注行变量和列变量的相互关系；或者说，人们不仅需要了解变量之间的关系，样品之间的关系，

还需要了解变量与样品之间的对应关系,这是主成分方法和因子分析等方法所不能说明的。这时候,对应分析就派上了用场。

因此,对二值行动者—事件的测量方法可采用对应分析(correspondence analysis)(Tools→Scaling/Decomposition→Correspondence)。对应分析很类似于潜类分析(Latent Class Analysis),其运算基础是多变量二值列联表分析(multi-variate binary cross-tabulations),有关它的分布假设更适用于二值数据。

对应分析是通过二维交互表的频数分布来确定"行动者"和"事项"及其类别之间的关系的,把这种关系反映在二维或三维的分布图中。每个变量的类别差异是通过图中的分值距离表现出来的,但是这个距离是加权的距离,在加权过程中考虑到卡方值之差异。因此,对应分析的基础是将卡方值转变为可度量的距离。距离越近,两个类别在图上的距离越近。

(二) 对应分析的应用

为了展示对应分析的应用,我们需要首先对数据进行二值化处理。如果分析目的主要是为了找到网络中的分派现象(如是否分为文理),就可以创建两类数据,一类要根据参与文科小组的情况,一类根据理科小组的情况(当然这需要事先知道有关文理科的情况)。从而分别针对这两个数据进行对应分析。

表 11.18 展示了利用对应分析法得到的小组在联合行动者—事件空间的三个维度上的负载值。

表 11.18 小组在联合行动者—事件空间的三个维度上的负载值

CORRESPONDENCE ANALYSIS

Input dataset: "H:\UCINET Data\DataFiles\affiliation.14.7
Method: Principal Coordinates
Matrix rank is 6
SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO	PRE	CUM PRE
1:	0.812	32.2	32.2	1.371	0.470	0.470
2:	0.592	23.5	55.7	1.223	0.250	0.720
3:	0.484	19.2	74.9	1.442	0.167	0.887
4:	0.336	13.3	88.2	1.834	0.080	0.967
5:	0.183	7.3	95.5	1.615	0.024	0.991
6:	0.113	4.5	100.0	1.615	0.009	1.000
	2.521	100.0				

Column Scores

		1	2	3
1	maths	−1.052	0.170	−0.117
2	politics	0.952	0.429	0.484
3	english	0.444	−0.046	−0.838
4	geography	−0.809	−0.503	0.701
5	chemistry	0.776	−1.495	−0.130
6	speech	0.574	0.524	0.136
7	biology	−0.981	0.299	−0.389

在对应分析中，给出的奇异值数目即维度数，它是两个变量（行数或列数）的最小类别数减去 1。例如，在我们的数据中，两个变量（学生和小组）的最小类别是 7（即小组数），因此，维度数为 6。从计算结果上看，PRE 表示“减少误差比例”，代表的是每个因子的解释力。可见，第一个因子的解释力为 0.47，累积到第三个因子的总解释力达到 0.887。可见，可以把因子数目确定为 3 比较恰当。每个因子的含义要结合每个学习小组在三个因子上的负载值来解释。负载值如 Column Scores 下面的数据所示。

在这个分析结果中，因子的含义不很明确，比较难以解释。这里需要指出的是，不同的解释者给出的解释可能完全不同，并且也不存在一个统一的解释。

总的来说，由于这些数据并没有反映明显的分派性，只反映了参与性，因此我们不能期望所得到的发现能够对应于上文讨论的结果。然而，我们确实看到了，这种方法也可用来对各个事件在各个维度上的负载值进行定位，从而能够把握住行动者和事件二者的方差。表 11.19 展示了各个行动者在三个因子上的负载值。

表 11.19　学生在联合行动者—事件空间的三个维度上的负载值

Row Scores

		1	2	3
1	A	−1.146	−0.281	0.603
2	B	0.809	0.510	−0.150
3	C	−0.738	−0.034	−0.332
…	…	… …	… …	… …
13	M	−0.312	0.399	−0.624
14	N	0.845	−0.248	−0.180

这里对因子的解释也不明确，因此，对每个学生的因子负载值的解释也比较

困难。因此,这里也不给出具体的解释。本部分分析的目的主要是为了展示其计算过程及结果。

如果希望区分出一些有意义的模式,那么可视化处理通常是较好的方法。图 11.12 展示了行动者和事件在联合对应分析空间中的前两个维度上的散点图。

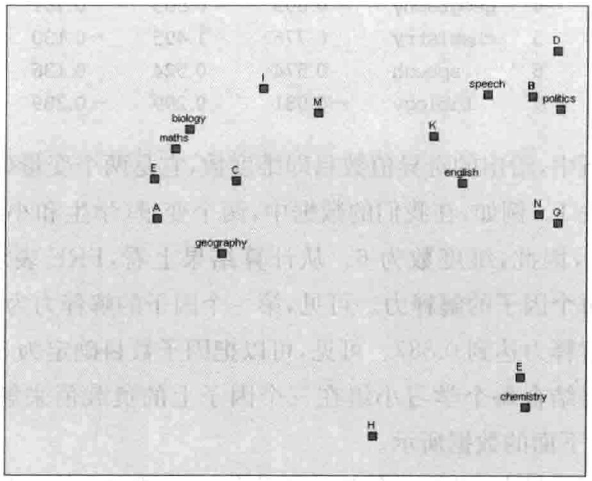


图 11.12 对应分析的二维图

右下方包含一个有意义的行动者和事件的聚类,表达了如何解释对应分析的结果。在左上方,我们看到 biology 和 maths 距离最近,在右上方,politics 和 speech 也最近。这些结果表明,小组分组比较明显,学生分组不太明显。

第三节 2-模数据的核心—边缘分析和分派分析

通常情况下,2-模网络中的行动者和事件都是"共在"(co-presence)的,发生阵一般是二值矩阵。在这种情况下,上述讨论的计量方法都是可以利用的,但是我们应该对计算的结果保持谨慎。这是因为,各种维度分析法都是在相似性/距离(similarity/distance)矩阵的基础上计算出来的。对于二值数据来说,诸如相关系数这样的测度可能引起误导(特别是在分析 2-模网络数据的时候更是如此)。当数据比较稀松的时候,甚至对应分析法也可能引起歧义。

在这种情况下,可选用块模型法(block modeling)来分析。块模型直接针对二值发生矩阵进行分析,它对行和列进行置换,找到尽可能与理想的像接近的结

果。这种研究用不着假定数据服从什么分布，而这些假设在通常的量表分析(scaling analysis)中都要有。针对1-模网络的块模型分析已经在上面的章节中有所介绍。下面介绍如何对2-模网络数据进行块模型分析，具体包括核心—边缘分析和分派分析。

一、2-模数据的核心—边缘分析

核心—边缘结构是一种理想型的结构模式，它把行和列都分为两类。在主对角线上的块是核心，是密度高的块，主对角线上的另外一块就是边缘，是密度低的块。核心—边缘模型不太关注非对角线上的块的关系密度。

当利用核心—边缘模型分析1-模数据的时候(这要利用UCINET中Network→Core/Periphery算法，参见刘军，2004：第八章)，该模型将努力区分出密度较高的一系列行动者(核心)，还要区分出密度较低的一系列行动者(边缘)。核心行动者在与边缘行动者的交换关系中处于优势地位。

表11.20 针对学生—小组数据进行类别性的核心—边缘模型分析结果

```
2-MODE CATEGORICAL CORE/PERIPHERY MODEL
------------------------------------------------------------------------
Input dataset:                   "H:\Ucinet Data\DataFiles\affiliation.

Starting fitness: 0.461
Final fitness: 0.743

Blocked Adjacency Matrix

                  2 1 1 1 1 1 2           1 1 1               1 1
                  1 7 8 9 5 6 0   2 4 3 1 2 3 4 5 6 7 8 9 0 1
                  b e g c m p s   B D C A L M N E F G H I J K
                 -------------------------------------------------
  3           C | 1 1 1   1     |                             |
 14           N |   1   1   1 1 |                             |
 13           M | 1 1     1   1 |                             |
                --------------------------------------------------
  1           A |     1   1     |                             |
  5           E |   1   1       |                             |
  2           B |   1       1 1 |                             |
  7           G |       1   1 1 |                             |
  6           F | 1   1   1     |                             |
  9           I | 1       1   1 |                             |
 10           J |   1       1 1 |                             |
 11           K |     1     1 1 |                             |
 12           L | 1   1   1     |                             |
  8           H |     1 1       |                             |
  4           D |           1 1 |                             |
 15       maths |               |                             |
 16    politics |               |                             |
 17     english |               |                             |
 18   geography |               |                             |
 19   chemistry |               |                             |
 20      speech |               |                             |
 21     biology |               |                             |
                 -------------------------------------------------

Density matrix

          1     2
      ----- -----
  1   0.571 0.000
  2   0.244 0.000
```

与之类似,当利用核心—边缘模型分析 2-模网的时候,实际上也一样,也要力图找到与理想的像(idealized image)接近的结果,即在主对角线上找到高密度和低密度的块,但是其意义却是不同的。“核心”是由一系列行动者的分区构成的,这些行动者与事件分区中的每个事件紧密联系在一起;同时,事件的分区也与核心分区中的行动者紧密联系在一起。因此,“核心”是一个聚类,是经常“共现”(co-occurring)的行动者和事件的聚类。“边缘”一方面由一系列行动者构成的分区构成,这些行动者在同样一些事件上不“共现”,另一方面由一系列事件的分区构成,这些事件相互没有联络,因为它们不共享一些行动者。

在 UCINET 中,沿着 Network→2-Mode→Categorical Core/Periphery 算法,利用其中的数量方法(numerical methods)来搜索行动者和事件的分区,使之与理想的像尽量接近。该算法所使用的搜索方法是一类生成性的算法(genetic algorithm)。它生成的拟合优度测度(measure of goodness of fit)要利用“拟合值”(fitness score)来表示(0 意味着拟合度低,1 意味着完全拟合)。我们也可以通过考察密度矩阵来判断结果的好坏。如果块模型完全拟合,那么 1.1 块的密度将是 1, 2.2 块的密度就是 0。表 11.20 展示了针对学生—小组数据进行 2-模核心—边缘模型分析的结果。

在计算结果中包含两个拟合优度指数,初始拟合值为 0.461,最终拟合值为 0.743。各个块的密度信息也体现在密度矩阵中。从这个结果来看,尽管它与理想的块有差距,但是该数据的分块矩阵中已经展示出“核心”,它由 C, N, M 和 7 个小组构成,其余数据处于边缘地位。

二、2-模数据的分派分析

另外一类块模型方法就是“分派”(factions)分析。分派分析在“凝聚子群”一章中已经有所交代,其目的是根据关系多寡对网络中的行动者进行完备且互斥的分组,使群体内部的密度高,群体之间的密度低。在 UCINET 中,沿着 Networks→Subgroups→Factions 程序可分析 1-模数据,可针对指定的分派数量(如 4 派)进行分析。而对于 2-模数据来说,则要利用 Network→2-Mode→2-Mode Factions 算法来分析,但是要注意,它只能分出两派。

在对 1-模行动者数据进行分派分析的时候,我们要做的工作是区分出两个行动者聚类,这两个聚类相互联系紧密,因为他们参与同样的事件,但是他们与其他派以及事件的成员之间的联系很少。如果我们利用这种分派思想来分析

1-模事件数据,我们要分析的则是一些事件,它们由于拥有相同的参与者而被联络在一起。

Network→2-Mode→2-Mode Factions 程序利用同样的方法分析行动者—事件矩阵。在分析过程中,程序对行动者和事件的联合分组进行定位,使这些分组尽可能互斥。原则上说,存在的分派数量可能多于两个,但是该程序只给出了两组。表 11.21 展示了上述数据的 2-模分派块模型分析的结果。

表 11.21 2-模网络的核心—边缘分析结果

```
2-MODE CATEGORICAL CORE/PERIPHERY MODEL
------------------------------------------------------------------------

Input dataset:                  "H:\Ucinet Data\DataFiles\affiliation.14
Output row partition:            H:\Ucinet Data\DataFiles\rowfactionspart
Output column partition:         H:\Ucinet Data\DataFiles\colfactionspart

Starting fitness: 0.194
Final fitness: 0.361
Correlation to ideal: 0.361

Blocked Adjacency Matrix

                1 1 2 1               1 1 1 1 1       1   1 2
                6 7 0 9   3 1 2 8 4 5 1 2 3 4 5 6 7 8 9 0 1
                p e s c   C A B H D E K L M N m F G g I J b
               ------------------------------------------------
11          K | 1   1   |                           1       |
 2          B | 1 1 1   |                                   |
13          M |   1 1   |                     1           1 |
 4          D | 1   1   |                                   |
 5          E |   1   1 |                                   |
16   politics |         |                                   |
 7          G | 1   1 1 |                                   |
 8          H |       1 |                           1       |
 9          I |     1   |                     1           1 |
10          J | 1 1 1   |                                   |
21    biology |         |                                   |
17    english |         |                                   |
18  geography |         |                                   |
14          N | 1 1 1 1 |                                   |
15      maths |         |                                   |
20     speech |         |                                   |
19  chemistry |         |                                   |
              -------------------------------------------------
 6          F |         |                     1     1     1 |
12          L |         |                     1     1     1 |
 3          C |   1     |                     1     1     1 |
 1          A |         |                     1     1       |
               ------------------------------------------------

Density matrix

          1     2
      ----- -----
 1    0.359 0.022
 2    0.063 0.172
```

计算结果显示,该网络分为两组“politics, english, speech, chemistry”和除了“F, L, C, A”以外的所有其他学生以及全部七个学习小组构成了网络的核心,其余关系构成网络的边缘。

计算结果提供了两个拟合优度指数:初始拟合值为 0.194,它是观察值(0 或 1)和“应该”出现在每块中的值之间的相关系数。最终拟合值为 0.361,它是重新定位后的元素(0 或 1)和“应该”出现在每个块中的值之间的相关系数。各个块的密度也告诉我们拟合优度的信息。对于分派分析来说,理想的模式应

该是在对角线上有1-块(即群体内关联紧密),非对角线上的值为0-块(群体之间的联系少)。

分为两派的模型的拟合结果没有进行核心—边缘模型分析的结果好(前者的拟合值为0.361,后者为0.743)。这说明,通过分派分析得到的数据的“像”(image)(即分为两个独立的,并且基本上联系较少的事件—行动者空间)并不像拥有一个高密度的“像”那样与实际数据拟合得好。

通过这种分派分析得到的两“块”本身也不是特别令人满意,因为第一块把多数行动者和全部的小组放在一派当中(密度为0.359),第二块的密度为0.172。

本章小结

2-模网络数据(隶属关系数据)可以使我们获得宏观—微观或者结构—能动关系的洞见。利用各种方法(矩阵法、图表法、计量法、中心性分析和块模型法等)来探讨2-模网络数据的模式。这种研究有助于我们描述、理解为什么行动者和事件如其所是地那样行动。因子分析(最好适用于多值数据)主要用来找到行动者—事件空间背后的维度,然后对行动者和事件都进行画图分析。在寻找这些抽象的维度“背后的逻辑”或者“潜在结构”方面,这些方法尤其有用。通过关注抽象的潜在维度,有助于区分出“共现”的行动者和事件的分组。

另一类方法是块模型。这些方法的目标是评价实际的行动者—事件隶属关系模式与有关行动者—事件的联合空间的理想型的模式(如核心—边缘模式或者分派模式)进行对比。就行动者—事件的隶属关系可以根据这种方式来理解而言,块模型也允许我们区分行动者所在的组别,也可以区分行动者所在的事件。

2-模数据分析不限于个体行动者及其所属的关系。用来分析2-模数据的工具也可用来分析认知性的社会结构数据,目的是根据依据网络理解的相似性来对理解者(行动者)进行分类。任何层次上的分析单位都可以看成是2-模问题。例如,多名读者和一些学术杂志之间的关系;大学生生源与学习能力之间的关系;学生的家庭背景与参与学校活动之间的关系等。在这个意义上可以说,2-模网络研究具有广泛的应用性。

第十二章

关系数据的假设检验*

在常规的统计分析如多元回归分析中，前提条件之一是要求多个自变量之间相对独立，不能高度线性相关，否则会出现“多重共线性”，它会引发一些问题。例如，在完全共线性的条件下，将得不到参数的估计量；在近似共线性情况下利用最小二乘法（OLS）得到的参数估计量无效，参数估计值不稳定（即根据不同样本计算出来的参数相差很大）。多重共线性使参数估计值的方差和标准差增大，变量的显著性检验失去意义，模型的预测功能失效。

在现实生活中，人们往往认为亲属之间更容易相互帮助。例如，“亲属”更愿意提供借款。“亲属”是一种关系，“借款关系”是另外一种关系，上一命题实际上论述的是这两类关系之间存在何种关系。现在的问题是，这两种“关系”之间的关系是否成立？这就需要对此进行假设检验。如果利用常规统计分析方法发现二者之间确实有关系，那么此时的相关性可能是虚假的，这一方面可能是由于村民之间在地理位置上的相近导致的；另一方面，这种研究本身就具有重大缺陷，即“亲属关系”和“往来关系”本身就有相关性。因此，我们不能利用常规的统计检验方法检验两类关系之间是否存在相关性。这时候，UCINET中的很多基于置换的检验方法（包括QAP）就派上了用场，它们可以对“关系”之间的关系进行假设检验。

在社会网络分析中，对关系命题的检验主要包括如下三类。第一类是点层次属性数据（node-level）的假设检验，其中包括回归分析（regression）、t检验（t-Tests）和方差分析（anova）。第二类是点—关系混合（mixed dyadic/nodal）层次的假设检验。它包括两类，一类是对“点的连续性数据”与“点与点之间的‘关系’数据”之间的关系进行检验，另一类是检验“点的离散性数据”与“点与点之间‘关系’数据”之间的关系。第三类是关系—关系层次（dyadic QAP）的假设检验，又分为三类。第一类是矩阵相关分析（QAP correlation），即对“两种关系数据”之间的关系进行检验；第二类是矩阵关系列联表分析（QAP relational crosstabs），

---

* 本章内容主要参考了汉尼曼（Hanneman，2008：第18讲）。

最后一类是矩阵的回归分析(QAP regression),它对作为因变量的一个关系矩阵与作为自变量的多个关系矩阵进行回归分析。下面分别介绍其应用。

第一节　点的属性层次的假设检验

所谓点层次(node-level)的假设检验,指检验的假设所涉及的变量都是关于点的属性(如性别、级别、职位等)层次变量。在 UCINET 中,这类检验又分为回归分析(regression)、t 检验(t-Tests)和方差分析(anova)三类,下面分别介绍。

一、回归分析

在社会网络数据中,如果已知行动者在定距层次上的属性变量,同时还知道其他定距自变量,要想研究它们之间的关系,就可以利用多元回归分析方法描述关系性质,预测关系走向。在 UCINET 中,可利用的程序是 Tools→Testing Hypotheses→Node-level→Regression。该程序根据常规最小二乘法计算基本的多元线性回归统计量,并估计其标准误和显著性水平。需要强调的是,这里的回归计算过程利用的方法是随机置换法(random permutations method),用这种方法来构建确定系数 R 的平方和回归系数的抽样分布。计算过程分为两步。第一步,对一个因变量向量和自变量向量的对应值进行标准的多元回归分析,这一步与常规多元回归分析一样;第二步,该程序对因变量向量的各个值所在的行进行随机置换,然后重新进行回归分析,保存计算出来的确定系数(R-square)和回归系数。第二步需要重复进行几百、几千或者几万次(这取决于研究者的设定),目的是为了估计统计量的标准误。对于每个系数来说,该程序都会计算在第一步计算中得到的系数值的情况占总置换次数的比例。下面用一例说明 UCINET 中多元回归的计算过程。

汉尼曼教授在其网络讲义(Hanneman, 2008)中使用了一个 2-模数据,该数据描述了在 2000—2004 年间,少数巨额捐款单位(至少捐款 100 万美元)为了支持或反对加利福尼亚的一些公投提案(ballot initiatives)而捐出巨款的情况①,

① 在加州以及美国的许多州,市民可以直接对公投法案(initiatives)进行投票立法,而相关的单位给予捐款支持或者反对立法。感谢美国加州大学河边分校(University of California at Riverside)罗伯特·汉尼曼(Robert A. Hanneman)教授向笔者惠寄该数据的原始信息并提供指导。本章下面的部分内容也主要参考了汉尼曼教授的讲义,特此向他表示感谢。

其中涉及23家单位和44项提案(参见表12.1)。数据有两种形式:多值和二值。多值数据包含三个取值(−1,0,1),描述的是捐款单位与各项提案之间的关系。具体地说,如果一个行动者反对某个提案,就在对应的元素上编码为−1;不反对,则编码为0,支持则编码为1。二值数据描述的仅仅是一个捐款单位是否对每个提案有贡献,有贡献(即捐款,不管是支持还是反对)则编码为1,否则编码为0。这23个捐款单位以及部分提案的多值数据如表12.1所示(节选其中的10个提案)。

表12.1　加利福尼亚捐款单位与提案资料(节选)

		1 OR	2 OR	3 OR	4 OR	5 OR	6 OR	7 OR	8 OR	9 OR	10 OR
1	TeachersAssn	0	0	0	0	0	0	−1	1	−1	−1
2	Cahuallas	0	−1	0	0	0	0	0	0	0	0
3	Morongos	0	0	0	0	0	0	0	0	0	0
4	PacTel	0	0	0	0	0	0	0	0	0	0
5	Hastings	0	0	0	0	0	0	0	1	0	0
6	Walton	0	0	0	0	0	0	0	1	0	0
7	Dems	0	0	0	0	0	−1	−1	1	0	0
8	Engineers	0	0	0	0	0	0	0	0	0	0
9	Bing	0	0	0	0	0	0	0	0	−1	−1
10	Restaurants	0	−1	0	0	0	0	−1	0	0	0
11	ServiceWorkers	0	−1	1	1	1	0	−1	1	−1	−1
12	Perenchio	0	−1	0	0	0	0	0	0	−1	−1
13	Hospitals	0	0	0	0	0	0	0	0	0	0
14	SchoolEmp	0	0	0	0	0	−1	−1	1	0	1
15	Reiner	0	0	0	1	0	−1	0	0	−1	−1
16	Builders	0	0	0	0	0	0	−1	1	0	0
17	StateEmp	0	0	0	0	0	−1	−1	0	0	0
18	CFT	0	0	0	0	0	0	−1	0	−1	−1
19	Fisher	0	−1	0	0	0	0	−1	1	0	0
20	Republicans	1	0	0	0	0	1	−1	0	0	0
21	AFSCME	0	0	1	0	0	0	−1	0	0	0
22	Intel	0	0	1	0	0	0	0	1	0	0
23	Chevron	0	−1	1	0	1	0	−1	0	0	0

该数据矩阵清楚地展示了捐款单位与提案情况。例如,数据第一行,即"加州教师协会"(The California Teachers Association)捐款反对第7、第9、第10号公投提案,却捐款支持第8号提案。"英特尔"(Intel)公司(第22行)捐款支持第3、第8号提案,并没有捐款反对什么提案的情况。

在UCINET中,我们可以利用Data→Affiliations(2-Mode to 1-Mode)程序,将该数据转换为两个1-模数据。其中一个1-模网络是捐款单位—捐款单位网络,它表示任意两家捐款单位"共同参与"的提案数;另外一个1-模网是提案—提案网,它表示任何两个提案"共同拥有"的捐款单位数。数据具体的转换方法可参见本书第十一章。

下面研究的问题是,这些捐款单位的中心度(作为因变量)与哪些自变量有关系?由于中心度指标有多种,为了明确起见,这里所说的中心度指"特征向量中心度"。在具体计算的时候,要针对上述捐款单位与捐款单位之间的网络(即

上述第一种1-模网,其元素被界定为捐款单位之间"参与共同"的提案数),利用Network→Centrality→Eigenvector程序,在"Input dataset"对话框中选择该1-模网络数据,点击OK后即可计算出每家单位的特征向量中心度,将计算结果保存并命名为EigenvectorCentrality,它是一个UCINET数据文件,在下面的多元回归中将用该数据作为因变量。

自变量有三个,涉及捐款单位的类别和捐款单位之间参与同类捐款事务的平均次数。为此,将这23家大宗捐款单位分为三类(资本家利益群体、工人利益群体、其他利益群体),并用UCINET中的一个列向量表示。也就是说,如果某家捐款单位属于"资本家群体",则将该单位所在列赋值为1。

在UCINET中的具体多元回归分析中,我们可以将计算出来的每家政治捐款单位的特征向量中心度(eigenvector centrality)这个属性变量看成是因变量。自变量有三个,一个是"资本家"(CAP),一个是"工人"(WORK)(这两个变量都是虚拟变量,表明每个捐款单位隶属于"资本家"群体,还是"工人"群体。参照群体是"其他",它构成了回归分析的截距项),第三个自变量是POSCOAL(Participants On the Same side of issues with other donors in COALition),它代表每个捐款单位与其他捐款单位参与同类捐款事务平均次数(负值表明与其他捐款单位参与不同的捐款事务)。这些变量的构造需要利用Data→Spreadsheets→Matrix数据表程序。前两个变量的构造很简单,按照各个单位的类别输入1或者0即可,变量具体的取值如表12.2所示。第三个自变量的构造要麻烦一点,需要计算1-模捐款单位—捐款单位网络矩阵每一行的均值(对角线上的值除外)。在UCINET中可以用矩阵操作方式达到这个目的。在UCINET中,点击Transform→Matrix Operations→Within dataset→Aggregations,在对话框"Input dataset"中找到所保存的1-模捐款单位—捐款单位网络矩阵,在"Arithmetic"中选择"Average",在

表12.2 多元回归分析中构建的三个自变量

	CAP	WORK	POSCOAL
Morongos	0	0	2.14
Republicans	1	0	-0.23
Restaurants	1	0	1
ServiceWorkers	0	1	3.64
Fisher	0	0	1.82
Perenchio	0	0	1.73
Hastings	0	0	2.73
Walton	1	0	1.86
Chevron	1	0	2.82
Reiner	0	0	2.91
Cahuallas	0	0	1.36
TeachersAssn	0	1	4.82
CFT	0	1	4.09
Engineers	0	0	0.86
Intel	1	0	1.64
Builders	1	0	3.91
PacTel	1	0	1.32
Hospitals	1	0	1.96
Bing	0	0	1.91
StateEmp	0	1	3
AFSCME	0	1	4.091
SchoolEmp	0	1	4.182
Dems	0	1	4

“Break-out results by”一项中选择“Rows”,在“(For sqr mats)Diagonal valid?”一项中选择“No”,点击 OK 后即可计算出每家单位与其他单位参与共同捐款事务的平均次数。构建完毕后,将数据保存,并与上述两个向量组合在一起,命名为 CA_3_group,其中包含的三列向量就作为多元回归分析的三个自变量。

在回归结果中,我们可以期望工人群体将比其他群体表现出较高的特征向量中心度,或者说,我们期待“左派”利益集团将表现出明显的高联络性,并且总的来讲其成员较多地与中心度高的成员相连。

下面研究的问题是,“特征向量中心度”这个因变量可以在多大程度上被三个自变量解释?或者说“工人”是否如数据表现的那样,拥有较高的“特征向量中心度”?“工人”集团的成员之间是否拥有较多的联系?这些联系与共同参与捐款事务有多大关系?要回答这些问题,就需要以“特征向量中心度”作为因变量,以上述“资本家”、“工人”和“参与同类捐款事务次数”为自变量进行多元回归分析。图 12.1 给出了多元回归的对话框。

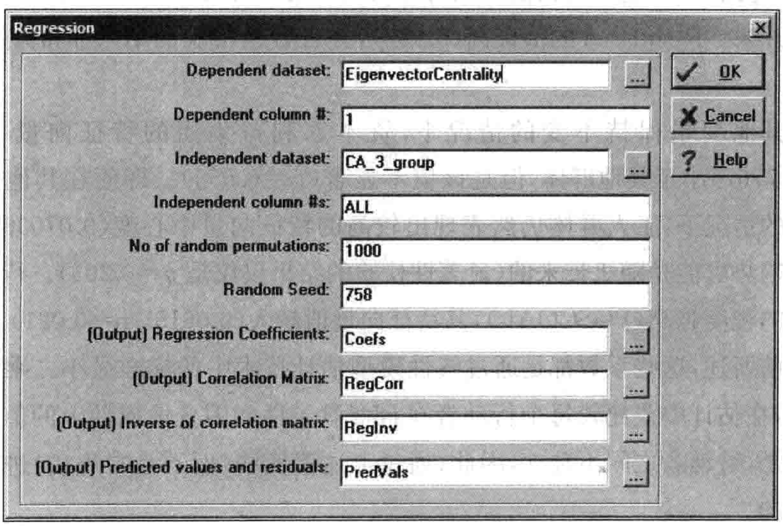


图 12.1　多元回归分析对话框

需要注意的是,全部自变量只需要用单独一个数据 CA_3_group 输入即可,因为该数据已经拥有三列,就相当于有三个自变量。计算结果如表 12.3 所示。

表 12.3 特征向量多元回归以及以置换为基础的显著性检验结果

CORRELATION MATRIX

	1	2	3	4
1	1.000	-0.483	-0.411	-0.480
2	-0.483	1.000	0.763	0.878
3	-0.411	0.763	1.000	0.970
4	-0.480	0.878	0.970	1.000

Determinant = 0.31843132

NOTE: All probabilities based on randomization tests.

MODEL FIT

R-square	Adjusted R-square	F Value	One-Tailed Probability
0.987	0.984	482.655	0.014

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	St'dized Coefficient	Proportion As Large	Proportion As Small	Proportion As Extreme
Intercept	0.003293	0.000000	1.000	0.000	1.000
CAP	-0.007767	-0.033723	0.555	0.445	0.911
WORK	0.075367	0.316154	0.203	0.797	0.423
POSCOAL	0.061454	0.714805	0.021	0.979	0.043

从相关矩阵可以看出工人群体(变量 3)与参与联络(变量 4)之间的相关系数极高,达到 0.97,这说明我们难以将参与联络的效应与成为工人利益群体的效应区分开来。R-squared 的值极高(0.987),并且置换检验的结果非常显著(p=0.014)。

在其他变量保持不变的情况下,资本家利益集团的特征向量中心度(−0.007767)小于其他群体,但是该值不显著(p=0.555)。即便在其他变量保持不变的情况下,工人群体仍然表现出较高的特征向量中心度(0.075367)。但是这种趋势可能是随机带来的(显著性检验的结果仅仅是 p=0.203)。行动者参与联络的程度越高(POSCOAL),其特征向量值越大(0.0615, p=0.021)。

如前所述,这些系数都是通过线性模型估计技术中的标准最小二乘法估计得到的,在估计中要比较每个行动者在自变量属性和因变量属性上的值。需要注意的是,行动者之间不独立,因此,通过上述置换法(而不是常规法)进行估计是必要的。

二、t 检验

在研究组织信息网的时候,如果根据某种理论或者经验提出一个假设,认为非政府组织要比政府组织更加倾向于分享信息,那么我们就可以通过比较这两类组织行动者的"平均点出度"来检验这个假设,这就要利用 t 检验方法(至于 t

检验的含义，请读者参考任意一本初级统计学教材即可)。

UCINET 自带一个数据 KNOKI，它是美国社会学家诺克(Knoke)教授收集到的多个组织之间信息交换的数据。该数据并非独立存在，而是保存在 UCINET 的 KNOKBUR 数据之中，该数据包含两个矩阵，其中的一个就是信息交换数据 KNOKI。为了利用 KNOKI，需要将 KNOKBUR 数据“开包”，这需要在 UCINET 中，点击 Data→Unpack 程序，选择 KNOKBUR 数据即可将它分解为两个数据，其中一个就是 KNOKI。然后计算 KNOKI 数据中各个点的点出度，即执行 Network→Centrality→Degree，对 KNOKI 分析即可，该程序会计算出各个点的度数中心度。将计算结果保存并命名为“FreemanDegree”，其中的第二列表示标准化的度数中心度(NrmDegree)，这一列才是我们要用到的连续变量。

为了检验这个假设，还需要对组织进行分类：非政府组织和政府组织。这需要利用 Data→Spreadsheets→Matrix 创建一个 UCINET 属性数据文件，可命名为“knokegovt”。该数据文件就是一个列向量数字，具体用两个数字表明两类组织的性质：“1”代表政府组织，“2”代表非政府组织。

下面开始执行一个简单的双样本 t 检验，目的是确定政府组织的度数均值是否小于非政府组织的度数均值。图 12.2 给出了执行这种分析的对话框。

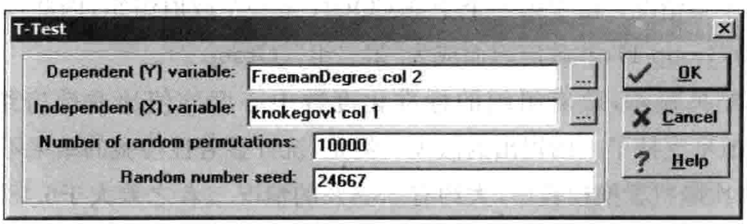


图 12.2　t 检验对话框

需要注意的是，在指定因变量名 FreemanDegree 的时候，由于标准化的中心度恰好位于该数据的第 2 列，因此，需要在因变量文件名 FreemanDegree 后面键入 col 2。与之类似，在图中“自变量”(Independent(X) variable)一项中，不但要指定数据文件名 knokegovt，还要具体指定该数据的哪一行或者哪一列，本例指定 knokegovt 文件数据的第一列(knokegovt col 1)，因为我们构造的就是第一列数据。

本例执行 10 000 次随机置换计算，算出两个均值之差的抽样分布(permu-

tation-based sampling distribution),计算出显著性水平。由于这是以置换为基础的假设检验,因此不需要常规的样本独立性假设和随机假设,其检验结果也比较适合于网络数据,如表 12.4 所示。

表 12.4 对 Knoke 政府和非政府组织的标准化度数中心度进行均值差检验的结果

```
Dependent variable:            FreemanDegree col 2
Independent variable:          knokegovt col 1
# of permutations:             10000
Random seed:                   24667

Basic statistics on each group.

                       1          2
                 Group 1    Group 2
               ---------  ---------
 1       Mean     75.000     68.519
 2    Std Dev      9.213     21.675
 3        Sum    300.000    411.111
 4   Variance     84.877    469.822
 5        SSQ  22839.506  30987.652
 6      MCSSQ    339.506   2818.930
 7   Euc Norm    151.127    176.033
 8    Minimum     66.667     33.333
 9    Maximum     88.889    100.000
10   N of Obs      4.000      6.000

SIGNIFICANCE TESTS

     Difference       ...One-Tailed Tests...        Two-Tailed
       in Means    Group 1 > 2      Group 2 > 1           Test
 ============== ============== ============== ==============
          6.481          0.334          0.750         0.6268
```

分析结果首先报告每个群体的描述统计量。群体的编码依据是包含自变量的文件中个案情况。由于第一个个案 COUN 是一个政府组织,因此,政府成为"第一组"(Group 1),非政府因而成为"第二组"(Group 2)。

计算结果显示,政府组织的标准化度数中心度均值比非政府组织的高 6.481,这似乎支持了上面提出的假设。但是,统计显著性检验的结果不容乐观。在 10 000 次随机置换检验中,大约有 33.4%的情况二者之差大于6.481。因此,如果我们认为观察的数据与研究假设是一致的,则犯错误的概率很大。也就是说,从统计意义上讲,政府组织和非政府组织在传递信息方面差别不大。

三、方差分析

t 检验只能比较"两个"群体均值之差,如果我们关注"多个"群体的均值是否相等,那么 t 检验就无能为力。这时候,方差分析就派上了用场。在 UCINET 中,Tools→Testing Hypotheses→Node-level→Anova 程序提供利用常规最小二乘法进行的方差分析,用它来估计多个群体均值之差是否相等。由于观察值不独立,因此,还是要利用随机重复的方法来估计标准误。如上所述,假设将 23

个捐款单位分为三类，用一个 UCINET 列向量表示，并且仍然保存在上述 CA_3_group 数据中的话，那么该矩阵第一列的值就变成三类：0 代表“其他”，“1”代表“资本家”，“2”代表“工人”。点击 Tools→Testing Hypotheses→Node-level→Anova，还是将“EigenvectorCentrality COL 1”作为因变量，将新的“CA_3_group COL 1”数据作为自变量，执行计算结果如表 12.5 所示。

表 12.5 对特征向量中心度进行一元方差分析及置换检验的结果

```
TOOLS>STATISTICS>ANOVA
------------------------------------------------------------------------------

Dependent variable:          EigenvectorCentrality Col 1
Independent variable:        CA_3_group Col 1
# of permutations:           5000
Random seed:                 20662

        ANALYSIS OF VARIANCE

         Source            DF             SSQ    F-Statistic   Significance
  ============== ============= ============== ============== ==============
       Treatment            2            0.21        34.4108         0.0002
           Error           20            0.06
           Total           22            0.28

R-Square/Eta-Square: 0.775
```

此前的计算已经表明，8 家“其他”捐款单位，8 家“资本家”捐款单位和 7 家“工人”捐款单位的特征向量中心度均值分别是 0.125，0.106 和 0.323。由方差分析结果可以看出，这些均值之差是显著的($F=34.4$，$p=0.0002$)。群体均值之差解释了各个捐款单位的特征向量中心度值总变异的 77.5%。需要注意的是，均值之差显著，但是并没有告诉读者具体哪两种行动者之间均值之差显著，例如“资本家”与“其他”的均值之差是否显著？该方差分析的结果不能给出这些内容。

最后需要补充的是，这里介绍的回归分析、t 检验和方差分析的结果都是“微观”意义上的，即在个体行动者层次上计算出来的结果。所使用的自变量和因变量可能是关系变量，也可能不是。也就是说，我们可能希望预测、检验关于行动者非关系属性(如收入)的假设，但是却利用关系性变量(中心度)与非关系性变量(如性别)的混合来检验。我们也可能希望利用关系变量和非关系变量的组合来预测行动者的关系属性(如中心度)。再强调一点：网络中的观测值不独立，因此，需要利用置换法直接估计出统计量的抽样分布，常规的统计软件不会给出正确答案。上述介绍的回归分析、t 检验和方差分析虽然在常规统计学中都有，但是在这里主要用它们来研究关系数据之间的关系。

第二节　点—关系混合层次的假设检验

所谓点—关系混合层次的假设检验，指的是假设检验涉及“点的属性”与“点与点之间的关系”。这类假设检验包括两类。一类是检验“点的连续性数据”与“‘关系’之间关系”之间的关系，另外一类是检验“离散性数据”与“‘关系”之间关系”之间的关系。下面分别论述。

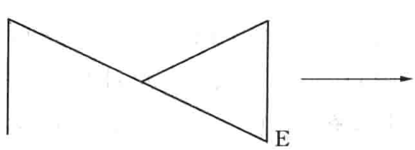
一、类别变量与关系变量之间关系的检验

（一）对两个群体内外关系模式的检验

有时候我们关注性别相同的人是否容易相互提供建议？也就是说，对于一个二元关系（如提供支持）邻接矩阵来说，它与一个单一变量（例如“性别”这样能够代表每个行动者属性的变量）之间是否相关？例如，“朋友关系”与“性别”是否有关？或者说朋友关系是否根据性别来分组，也就是说是否男孩倾向于与男孩建立朋友关系，女孩倾向于与女孩建立朋友关系？可采用两种办法来检验。

1. 利用 UCINET 中的“自相关”(AUTOCORRELATION)分析法。

这需要根据 UCINET 中的 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Categorical Attributes→Joint-Count 程序。该程序可以检验两个群体内部和群体之间的关系密度是否有别于随机分布的情形。具体地说，该程序对一个对称矩阵和二分向量之间的关系进行自相关随机检验(randomization test of autocorrelation)。需要注意，该程序只能针对分为两组(如男孩和女孩)的数据进行检验，检验的基础是计算组内和组间的关系数量，并且与一种随机的模型进行比较，如图 12.3 所示。



	A	B	C	D	E
A	—	1	0	0	0
B	1	—	1	0	0
C	0	1	—	1	1
D	0	0	1	—	1
E	0	0	1	1	—

图 12.3　五人之间的建议关系及其矩阵

假设上述五人图是“建议关系图”，其中每个人(A，B，C，D，E)的性别分别是男，女，男，男，女。将“男”编码为“1”，“女”编码为“0”，则可以构建性别列向量(1，0，1，1，0)。在 UCINET 中输入该列向量，并令其为 sex。然后计算该向量与“建议关系”矩阵之间的关系。点击 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Categorical Attributes→Joint-Count 之后，出现如下对话框，键入(或选出)输入事先已知的数据建议关系数据“advice”和已知的性别分区列向量 sex，如图 12.4 所示。

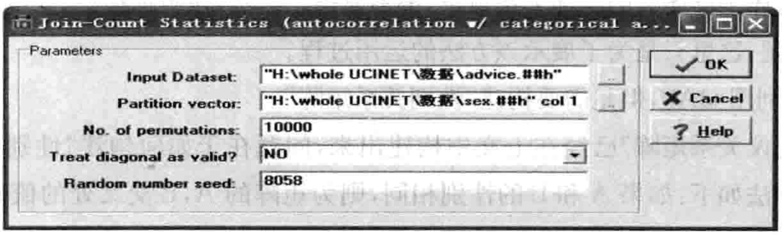


图 12.4　二方一点属性之间关系的假设检验对话框

点击 OK 后即计算出如下结果：

表 12.6　二方一点属性之间关系的假设检验结果

```
CATEGORICAL AUTOCORRELATION: JOIN-COUNT STATISTICS
------------------------------------------------------------------
Adjacency dataset:            H:\whole UCINET\数据\advice
Attribute:                    "H:\whole UCINET\数据\sex.##h" col 1
# of Permutations:            10000
Random seed:                  8058

Warning: Row Attribute vector has been recoded.
Here is a translation table:

  Old Code     New Code
  ========     ========
      0      =>      1
      1      =>      2

Number of iterations = 10000

                    1          2          3          4          5
             Expected   Observed  Differenc  P >= Diff  P <= Diff
            ---------  ---------  ---------  ---------  ---------
  1 1-1         0.500      0.000     -0.500      1.000      0.504
  2 1-2         3.000      4.000      1.000      0.304      1.000
  3 2-2         1.500      1.000     -0.500      1.000      0.605
```

可见，计算结果给出了实际的关系数量和期望的数量。第一行给出了群体 1 的内部关系数，即女孩和女孩之间的关系数；第二行给出了群体 1 和群体 2 之间的关系数，即女孩和男孩之间的关系数；第三行给出了群体 2，即男孩群体内部的关系数。其中，期望值(expected)指的是，如果各个 1 是随机地在群体内部和群体之间分布的话，所能出现的“1”的数量，它实际上等于全部可能出现的关

系除以 2。观察值(observed)就是实际出现的关系数量。男孩和男孩之间的观察到的关系数为 1 个,这个“1”对应着下面的“建议关系矩阵”中“第 3 行第 4 列”的“1”。“差量”(difference)就是把前者减去后者得到的数,即观察值减去期望值。P＞＝Diff 和 P＜＝Diff 这两项给出的是根据随机重排矩阵计算出来的差量大于(小于)或等于观察值的相对频次概率(relative frequency)。这些列的值可用来检验观察值的显著性。

从计算结果来看,无论是群体 1 的内部关系数量,群体 1 和群体 2 之间的关系数量,还是群体 2 的内部关系数量,都不显著。当然,该数据很小,并不能说明什么问题,这里只是为了展示该方法的运用过程。

2. 利用 QAP,但是需要构建“性别关系矩阵”。

“建议关系矩阵”已经在上文中构建出来,问题在于如何构建“性别关系矩阵”? 方法如下:如果 A 和 B 的性别相同,则为矩阵的 A,B 交叉处的值赋予 1,否则为 0。

假设 A=男,B=女,C=男,D=男,E=女。那么根据上述思想构建的“性别关系矩阵”应为(在建议关系矩阵中,C 行 D 列交叉值为“1”,它与表 12.6 中最后一行“3 2-2”和“Observed”一列交叉值“1”是对应的)以及前文给出的建议关系矩阵:

性别关系矩阵:

	A	B	C	D	E
A	—	0	1	1	0
B	0	—	0	0	1
C	1	0	—	1	0
D	1	0	1	—	0
E	0	1	0	0	—

;建议关系矩阵:

	A	B	C	D	E
A	—	1	0	0	0
B	1	—	1	0	0
C	0	1	—	1	1
D	0	0	1	—	1
E	0	0	1	1	—

然后,就可以利用 QAP(QAP 的具体含义见下一节)计算二者之间的相关关系了,具体计算步骤展示参见下一节,下面只展示计算的结果,如表 12.7 所示。

可见,“性别关系”与“建议关系”之间负相关,尽管这种关系在统计意义上不是显著的。读者完全可以根据自己的课题,探讨与本例相似的相关关系问题。

由上述两种方法计算结果可知,方法 1 给出组内和组间的检验结果,计算出显著性水平;方法 2 只给出组间检验,另外计算出相关系数和显著性水平。

表 12.7 QAP 检验结果

```
QAP CORRELATION
--------------------------------------------------------------------------------
Data Matrices:                advice.##h
                              samesex.##h
# of Permutations:            5000
Random seed:                  24322

QAP results for samesex.##h * advice.##h (5000 permutations)

                    Obs Value Significa Average Std Dev Minimum Maximum Prop>=0 Prop<=0
                    --------- --------- ------- ------- ------- ------- ------- -------
Pearson Correlation    -0.408     0.296   0.003   0.364  -0.408   0.816   1.000   0.296

QAP Statistics

QAP Correlations
             advice samese
             ------ -----
     advice  1.000  -0.408
    samesex -0.408   1.000

QAP P-Values
             advice samese
             ------ ------
     advice   0.000 0.296
    samesex   0.296 0.000
```

（二）对多个群体内外关系模式的检验：关系列联表分析

社会科学研究中有一个基本原理，即“社会分组”原理（谢宇，2006：9—28）。在现实生活的某种关系中，这种“物以类聚，人以群分”原理是否成立？以前文“大额捐款单位”数据为例，前文的分析已经发现，“非工人”（non-worker）的群体都没有表现出社会群体的聚类性，“工人”集团表现出凝聚性。

分析聚类性的方法不止一种。在 UCINET 中，关系列联表（Relational Contingency-Table analysis）分析程序可以对诸如此类的多个群体的关系模式提供更好的检验。该程序会对一个非对称的邻接矩阵与分组变量之间的关系进行自相关随机检验。例如，如果邻接矩阵是五个人之间的“友谊关系矩阵”，分类变量是性别，那么该程序会检验“朋友关系”是否因性别的不同而表现出不同的模式。该程序与标准的卡方检验类似，只不过该程序并不用卡方分布进行推断，而是根据随机化过程得到统计量的分布。下面仍然以“捐款单位数据”为例加以说明。

仍然将政治捐款单位分为其他（others）、资本家和工人三类。点击 Tools→Testing Hypotheses → Mixed Dyadic/Nodal → Categorical Attributes → Relational Contingency-Table analysis 程序后出现对话框，在“Input dataset”中键入事先转换的“捐款单位—捐款单位数据矩阵”，在属性变量“attribute”一项中填入已经构造好的“三类属性”数据。点击 OK 后，分析的结果如表 12.8 所示。

表 12.8　三个群体在密度方面的差异检验结果

```
 Old Code     New Code    Frequency
 ========     ========    =========
     0    =>      1           8
     1    =>      2           8
     2    =>      3           7

Number of ties: 87.000

Cross-classified Frequencies

          1  2  3
          0  1  2
         -- -- --
 1 0     12  9 14
 2 1      9 13 11
 3 2     14 11 28


Expected Values Under Model of Independence

              1      2      3
              0      1      2
          ----- ----- -----
 1 0      9.63 22.01 19.26
 2 1     22.01  9.63 19.26
 3 2     19.26 19.26  7.22


Observed/Expected

            1     2     3
            0     1     2
         ---- ---- ----
 1 0     1.25 0.41 0.73
 2 1     0.41 1.35 0.57
 3 2     0.73 0.57 3.88


Observed chisquare value = 74.217


Average permutation frequency table

           1      2      3
       ----- ----- -----
 1     15.03 16.14 14.10
 2     16.14 15.14 14.26
 3     14.10 14.26 12.33

Significance = 0.000200
Number of iterations = 10000
```

可见,程序将“其他”群体重新编码为“1”,将“资本家”群体重新编码为“2”,将“工人”群体重新编码为“3”。整个图包含 87 个关系,三类群体的交互分类频次如“Cross-classified Frequencies”一项所示。下面的表格给出的是在三种利益集团“独立”的情况下出现的期望值。从分析结果中可以看到,观察值不同于在“独立模型情况下的期望值”(Expected Values Under Model of Independence)。下一个表给出了观察值与期望值之比(Observed/Expected)。可见,对角线上的值(即群体内部关系)都大于 1,表现出同类性(homophily)。

检验结果给出的卡方统计量是 74.217。随后给出的是在 10 000 次随机置换中计算出来的平均关系频次,显著性水平 p=0.0002。由此得到的结论是,非交互模式不出现的概率极低,或者说三类捐款单位之间存在交互关系。

(三) 对多个群体内外关系模式的检验:方差密度模型

一个群体的成员可能愿意与自身群体的成员交往,也可能不愿意。现实情况如何? 除了上述模型之外,在 UCINET 中,Tools→Testing Hypotheses→

Mixed Dyadic/Nodal→Categorical Attributes→ANOVA Density Models 程序还提供了另外一些检验“相似性模型”的方法，下面分别介绍。这些检验方法根据的是每一块的密度，它类似于方差分析。具体的模型包括如下四类：(1)常量相似模型(Constant Homophily)，该模型的前提假设是，行动者所倾向的交往对象与自己的类别相同，即隶属于同类类别，并且假设所有的类别群体都有同样的交往趋势；(2)变量相似性(Variable Homophily)，与常量相似模型类似，但是它假定每一类群体拥有的交往趋势是可变的；(3)结构块模型(Structural Blockmodel)，这是一种应用最广泛的模型，它关注不同的群体是否拥有明显不同的交往模式。例如，女孩可能愿意与女孩交往(inbreeding)，而男孩也愿意与女孩交往(outbreeding)。除了这三个模型之外，还有核心—边缘模型-1(Core/Periphery 1)和核心—边缘模型-2(Core/Periphery 2)。下面分别介绍这些模型。

1. 结构块模型(Structural Blockmodel)法。

该方法可以检验不同群体的群体内外关系模式的差异，但是并不指出差异何在。以上文介绍的“捐款单位—捐款单位”数据为例。假设我们仅对“强联盟关系”的存在感兴趣，为此我们需要给出新的界定：如果两个行动者(即两家单位)共同参与四项及以上捐款事务，则界定二者之间的关系为“强”，并在对应的行和列的值赋予“1”，否则为“0”。要给出该数据矩阵，需要在 UCINET 中，点击 Transform→Dichotomize，在出现的对话框中，“Input dataset”项上选择前文给出的“捐款单位—捐款单位”1-模网络数据，在“Cut-Off Operator”项上选择“GT-Greater Than”，在“Cut-Off Value”项上将默认值“0”修改为“4”，将“Diagonal OK”修改为“No”，点击 OK 即可将该多值数据转换为二值数据。分类变量仍然是前文给出的三类列向量。

在 UCINET 中，点击 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Categorical Attributes→ANOVA Density Models，在对话框“Network or proximity matrix”一项中选择上述二值化处理的数据矩阵，在“Actor Attribute(s)”项中选择前文给出的三类属性列向量，在“Model”项中选择“Structural Blockmodel”，点击 OK 后得到计算结果。表 12.9 给出了三类群体(工人、资本家、其他)中强联盟关系(共同参与四项及以上事务)的模型拟合结果。

表 12.9 群体关系密度之差的结构块模型分析结果

```
Density Table

              1     2     3
              0     1     2
            ----- ----- -----
   1 0     0.143 0.141 0.250
   2 1     0.141 0.179 0.196
   3 2     0.250 0.196 1.000

Density table saved as dataset C:\Documents and S

Expected values saved as dataset C:\Documents and

Number of permutations performed: 5000

MODEL FIT

R-square Adj R-Sqr Probability    # of Obs
-------- --------- ----------- -----------
   0.276     0.266       0.000         506

REGRESSION COEFFICIENTS

              Un-stdized      Stdized
 Independent Coefficient Coefficient Significance
 ----------- ----------- ----------- ------------
   Intercept    1.000000    0.000000        0.000
         1-1   -0.857143   -0.618590        0.000
         1-2   -0.859375   -0.657102        0.000
         1-3   -0.750000   -0.541266        0.000
         2-1   -0.859375   -0.657102        0.000
         2-2   -0.821429   -0.592815        0.000
         2-3   -0.803571   -0.579928        0.000
         3-1   -0.750000   -0.541266        0.000
         3-2   -0.803571   -0.579928        0.000
```

首先给出实际数据的密度表。可见,“其他”群体的成员自身之间联系比较少(0.143),与“资本家”群体的联系也不多(0.141),但是与“工人”群体的联系多一些(0.250)。只有“工人”群体(即第三行第三列)内部成员之间的联系多。

随后给出的是回归模型。因变量是每一对行动者之间的关系,对代表 3×3 块模型表中每个元素的一系列虚拟变量作回归。在该回归模型中,最后一类(即第三行第三列)是参照类。从结果可以看出,块之间的差异解释了成对关系变异的 27.6%。如果两个行动者都隶属于“工人”群体(第 3-3 块),那么二者之间存在关系的概率是 1.000(即截距项)。如果两个行动者都隶属于“其他”群体(第 1-1 块),那么二者之间存在关系的概率比 1.000 小 0.857。

需要再次强调的是,如果利用常规的只针对独立数据的公式进行显著性检验,就不会得到正确的答案。相反,如果针对关系存在与否进行 5 000 次随机置换,从得到的模型抽样分布中估计出恰当的标准误。

2. 常量相似法(Constant Homophily)。

在 Tools → Testing Hypotheses → Mixed Dyadic/Nodal → Categorical

Attributes→ANOVA Density Models 中,还存在一个对群体之差进行检验的更加严格的方式,即常量相似法模型。该模型假定全部群体都有一种建立群体内关系的倾向,并且所有群体的关系强度都相同。利用这种模型检验的结果表明模型的拟合效果不好,具体检验过程略。

3. 变量相似法(Variable Homophily)。

通过上述模型的分析可以看出,只有"工人"群体内存在较强的关系。我们还可以利用第三类模型,即变量相似法来检验这个命题,该模型检验的是每个对角线内的关系(即每个群体内部的关系)是否有别于对角线外的关系。在 UCINET 中,点击 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Categorical Attributes→ANOVA Density Models,再点击"Model",在下拉式菜单中选择 Variable Homophily,对上述数据进行分析,得到的结果如表 12.10 所示。

表 12.10 群体关系密度之差的变量相似块模型分析结果

MODEL FIT

R-square	Adj R-Sqr	Probability	# of Obs
0.269	0.266	0.000	506

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	Stdized Coefficient	Significance
Intercept	0.193182	0.000000	0.999
Group 1	-0.050325	-0.036319	0.393
Group 2	-0.014610	-0.010544	0.487
Group 3	0.806818	0.512045	0.000

该模型的拟合效果比第二个模型好(R-square=0.269, p<0.000)。这里的截距项表明,不同群体的两个行动者之间存在关系的概率是 0.193。群体 1(即"其他"群体)中的行动者之间存在关系的概率比 0.193 小 0.05(但是差异不显著)。群体 2(即"资本家"群体)内部行动者之间存在关系的概率仅仅比0.193小0.0146(差异也不显著)。然而,群体 3(即"工人"群体)内部行动者之间存在关系的概率却比 0.193 大 0.8068(这次差异却是显著的)。

4. 核心—边缘模型法。

从上面的分析中可以看出,仍然只有群体 3 表现出群体内关系。这引导我们考虑利用核心—边缘块模型进行拟合检验。核心—边缘有两类:强模型和弱模型。强模型(Core-periphery 1)假定存在一个高度结构化的核心,即在群体内部存在许多关系,但是其他关系(边缘群体的成员之间,或者核心群体与边缘群体之间的关系)较少。表 12.11 给出了利用该模型拟合加州捐款单位数据的分

析结果。

表 12.11 对加州捐款单位数据进行强核心—边缘块模型分析的结果

MODEL FIT

R-square	Adj R-Sqr	Probability	# of Obs
0.008	0.008	0.394	506

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	Stdized Coefficient	Significance
Intercept	0.266667	0.000000	0.245
Core	-0.123810	-0.089352	0.245

显然,该模型并没有很好地描述群体内外关系模式。R-square 非常低(0.008),并且有 39.4%的随机置换数据会达到 R-square 这么低的结果。回归系数也不显著,“边缘”的密度(或者两个随机行动者之间存在关系的概率)是 0.27,“核心”的密度比 0.27 小 0.12。

弱模型(Core-Periphery 2)给出一种放宽条件的块模型。该模型仍然假定存在一个高度联系的核心,但是允许核心与边缘群体的成员之间存在关系。边缘群体成员之间的关系在极端情况下不存在关系。表 12.12 给出的是利用该模型拟合加州捐款单位数据的结果。

表 12.12 对加州捐款单位数据进行弱核心—边缘块模型分析结果

MODEL FIT

R-square	Adj R-Sqr	Probability	# of Obs
0.037	0.037	0.109	506

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	Stdized Coefficient	Significance
Intercept	0.352381	0.000000	0.068
Non-Periphery	-0.169949	-0.192629	0.068

可见,尽管该模型的拟合程度好一些(R-square=0.037),但是 R 平方仍然很低,并且有 11%的随机置换数据会达到 R-square 这么低的结果。截距项的密度(可界定为“边缘”)较高,大约是全部关系的 35%。在“非边缘”的任何两个单位之间存在关系的概率比 0.35 小 0.17。

二、连续变量与关系变量之间关系的检验

如果行动者的“类型”相同,那么他们在网络中是否“接近”?尽管该问题涉

及上文介绍的离散变量的"物以类聚"模型，这种思想也可以推广到连续变量。

在 UCINET 中，点击 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Continuous Attributes→Moran/Geary statistics，该程序提供两种指数来检验行动者在"定距测量层次的属性变量"和"二者之间的网络距离"之间是否存在自相关(autocorrelation)，这两种指数是 Moran's I 和 Geary's C，它们最初来自地理学，随后出现在空间统计学，测量的是任何两个地点在地理特征上的相似性是否与二者之间的空间距离有关系。Geary's C 指数对成对数据之间进行比较(paired comparisons)，而 Moran's I 系数关注的是协变异(covariations)。

Moran's I 指数是莫兰(Moran)于 1948 年提出来的，它反映的是空间上邻接的单元属性值的相似程度，一般用来检验空间中是否存在聚类，其计算公式为：

$$I=\left(n\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}(x_i-\bar{x})(x_j-\bar{x})\right)\Big/\left(\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$$

其中，x_i 是区域 i 的属性值，C 是空间权重矩阵，C_{ij} 代表空间单元 i 和 j 之间的影响程度，$\bar{x}$ 是各个 x 的均值。Moran's I 指数的取值范围在－1 和 1 之间，越接近－1，则表示单元之间的差异越大，越接近 1，则表示单元之间的关系越密切，接近 0 则表示单元之间无关。Moran's I 统计量的构造类似常规相关系数的构造，相当于两个变量的协方差除以方差。具体还请读者参见相关的空间统计学教材。

Greay's C 指数公式为(参见 http://www.lpc.uottawa.ca/publications/moransi/moran.htm)：

$$C=\left((n-1)\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\ (x_i-x_j)^2\right)\Big/2\left(\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\right)\left(\sum_{i=1}^{n}\ (x_i-x)^2\right)$$

它反映的是邻接空间值的成对比较。G 系数的取值范围大都在 1 和 2 之间。如果 G 介于 0 和 1 之间，则表明正的空间自相关；G 介于 1 和 2 之间，则表明负的空间自相关(这一点似乎不容易理解)。然而有时候计算出来的 G 也大于 2 (Griffith, 1987)。

现在我们把关注点从"空间"转向现实世界网络。假设研究的问题是，"距离近"的政治利益群体在捐款行为上是否也类似？如果经常联络的利益群体会相互影响，我们就可以利用大宗捐款(500 万元)者的信息来展示这种自相关问题的分析过程。

首先，创建属性文件数据，用一列数字表示每个点的贡献值，即捐款额。其

次,创建一个矩阵数据,用来“描述”每个点的“接近性”(closeness)。创建矩阵的方法比较多。一种是利用邻接矩阵。如果两个捐款单位至少共同参与了四项捐款事务,则令二者为邻接关系(可以通过对1-模数据进行二值化处理得到此类邻接关系数据)。我们也可以利用行动者之间关系的连续测度来测量“接近性”(closeness)。例如,可以利用捐款单位的捐款表现(取值从负数到正数,负数意味着两家单位参与过相反的提案捐款,正数意味着二者共同参与过同类提案捐款)来表示相似性指标(scale of the similarity)。我们也可以利用其他指标表示行动者的网络接近度(例如捷径距离的倒数)。总之,任何描述行动者接近程度的“相似性”(proximity)矩阵都是可以利用的,其中的一些矩阵可以利用 Tools→Similarities 和 Tools→Dissimilarities & Distances 计算出来。

表12.13和表12.14展示了 Tools→Testing Hypotheses→Mixed Dyadic/Nodal→Continuous Attributes→Moran/Geary statistics 的分析结果。这里利用邻接性作为网络距离,测量了行动者捐款额的自相关性。关注的问题很简单:在网络中邻接的行动者是否倾向于捐出相同数目的资金?这里展示了两个统计量和相关的信息。表12.13给出的是 Moran's I 统计量,表12.14给出的是 Geary's C 统计量。

表 12.13 邻接的政治捐款单位在捐款额方面的 Moran 自相关分析结果

```
Proximities:                C:\Documents and Settings
Attribute(s):               CA_3_group Col 1
Method:                     Moran
# of Permutations:          1000
Random seed:                611

NOTE: Larger values indicate positive autocorrelation.
      A value of -0.045 indicates perfect independence.

        Autocorrelation:       -0.119
           Significance:        0.174

    Permutation average:       -0.043
         Standard error:        0.073
    Proportion as large:        0.826
    Proportion as small:        0.174
```

我们已经指出,Moran's I 自相关统计量(最初用来测量空间自相关,这里用来测量网络自相关)的取值范围从−1(完全负相关)到0(无关),再到1(完全正相关)。这里的分析结果为−0.119,表明参与联络的两个行动者在其贡献水平上的表现要比随机的两个行动者的表现差异更大,但是该结果在统计意义上不显著。

这里还要应用置换检验,从而产生一个抽样分布。在每一次置换中,花费值被随机赋予行动者,计算出 Moran's I 统计量。在这些随机试验中,观测到的

Moran's I统计量均值是−0.043，标准差是0.073。观测值(−0.119)与随机情况下的期望值(−0.043)之差相对于抽样变异来说较小。实际上，在所有的随机样本中，17.4%的样本至少表现出如此大的相关性，远远高于常规的5%的可接受错误比例。

计算出来的Geary相关测度以及相关信息如表12.14所示。

表 12.14　邻接的政治捐款单位在捐款额方面的 Geary 自相关分析结果

```
Proximities:                     C:\Documents and Settings
Attribute(s):                    CA_3_group Col 1
Method:                          Geary
# of Permutations:               1000
Random seed:                     592

NOTE: Smaller values indicate positive autocorrelation.
      A value of 1.0 indicates perfect independence.

           Autocorrelation:        2.137
             Significance:         0.026

       Permutation average:        1.004
            Standard error:        0.613
       Proportion as large:        0.026
       Proportion as small:        0.974
```

前文已经指出，如果两个变量无关联，则Geary's C统计量等于1。若其值小于1，则意味着正相关，大于1意味着负相关。计算出来的值为2.137，表明负自相关，与Moran's I统计量的计算结果一样。但是与Moran统计量不同的是，Geary's C统计量指出，我们的结果与1 000次随机试验得到的均值(1.004)之差是显著的(p＝0.026)。

两种指标似乎测量的不是一类问题。Geary's C统计量的计算过程要考察每一对行动者取值之差，用他们的临界值进行加权。而Moran统计量的构建要考察每个行动者的取值与均值之差，然后对叉积(cross-products)进行加权。这两种统计量的差别意味着，Geary统计量比较关注每一对行动者的不同成员之间的差别，这是一种"局部"(local)差异，而Moran统计量关注的是，相同的或者不同的一对行动者相对于总体均值来讲是如何不同的，这是一种"总体"(global)差异。因此，在空间统计学中，有学者认为Geary统计量有时候对"局部"差异的敏感度比对"整体"差异的敏感度高(这从其公式中可以看出来)。

如果数据表现出较大的变异性，呈非正态分布，那么就"网络邻接性"对"属性相似性"的影响这一点来说，上述两类测度似乎表明不同的结果。需要强调的是，到底应该用哪种指数？这里并无对错之分，通常情况下最好将两个指标都计算出来，除非我们有明确的理论指明需要其中的哪种指标。

表12.15和表12.16是对上述例子的重复，但是有一点不同。在这两个例子

中,我们用每一对行动者共同参与的捐款事务数目来测量网络中两个行动者的接近性(当然也可以利用其他指标,如捷径距离等来测量接近性)。

表 12.15 对网络中接近的政治捐款单位的捐款额进行 Moran 自相关分析的结果

```
Proximities:              C:\Documents and Settings
Attribute(s):             CA_3_group Col 1
Method:                   Moran
# of Permutations:        1000
Random seed:              81

NOTE: Larger values indicate positive autocorrelation.
      A value of -0.045 indicates perfect independence.

         Autocorrelation:       -0.145
            Significance:        0.018

     Permutation average:       -0.047
          Standard error:        0.049
     Proportion as large:        0.982
     Proportion as small:        0.018
```

利用这种连续性的网络接近性测度(而不是邻接性)后,我们可能期望有较强的相关关系。现在,Moran 测度值是−0.145(与此前的−0.119 相比减少了),显著性水平为 p= 0.018。"接近"的行动者之间有一种较小但显著的趋势,即他们要比随机选出来的两个行动者更容易捐出不同数目的钱,这是一种负的网络自相关。

表 12.16 对网络中接近的政治捐款单位的捐款额进行 Geary 自相关分析结果

```
Proximities:              C:\Documents and Settings
Attribute(s):             CA_3_group Col 1
Method:                   Geary
# of Permutations:        1000
Random seed:              236

NOTE: Smaller values indicate positive autocorrelation.
      A value of 1.0 indicates perfect independence.

         Autocorrelation:        1.836
            Significance:        0.009

     Permutation average:        1.006
          Standard error:        0.438
     Proportion as large:        0.009
     Proportion as small:        0.991
```

利用这种连续性的网络接近性测度计算后,Geary's C 指数稍微减少一些(从 2.137 减少到 1.836)。这个结果仍然表明一种负的网络自相关,并且具有统计显著性。也就是说,如果在网络距离和费用方面无关的话,那么这种负相关不太可能随机地发生。

第三节 关系—关系层次的假设检验

在 UCINET 中,第三类假设检验是关于两类"关系"之间关系的假设检验(dyadic QAP)。它又包括三类:(1)矩阵相关分析(QAP correlation);(2)矩阵

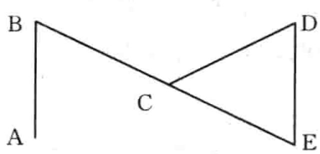
关系列联表分析(QAP relational crosstabs);(3)矩阵回归分析(QAP regression)。这种检验涉及一个重要方法,即 QAP。下面首先介绍 QAP 的含义,然后分别介绍其应用。

一、QAP 的含义

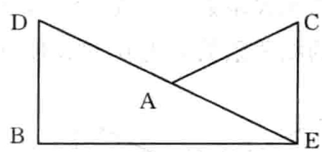
社会关系是社会学研究的核心内容之一。在研究关系之间关系(例如,“亲属关系”是否带来“经济支持关系”)的时候,需要特定的方法,QAP 就是方法之一。

QAP(quadratic assignment procedure,二次指派程序)是一种对两个方阵中各个元素的相似性进行比较的方法,即它对方阵的各个元素进行比较,给出两个矩阵之间的相关系数,同时对系数进行非参数检验,它以对矩阵数据的置换为基础。其原理如下(参见 Everett, 2002)。

假设用两个矩阵分别表示 5 个人(A, B, C, D, E)之间的“朋友关系”和“建议关系”(假设这两种关系都是对称的),如图 12.5 所示(1 代表朋友,0 表示不是朋友)。每两个人之间的关系都对应着关系矩阵中的元素,矩阵中的各行和各列对应着 5 个人。一般情况下,矩阵的元素只有两个:1 和 0。从(1)中可见,A 和 B 是朋友,B 和 D 不是朋友,C 和 E 是朋友,其他情况都可从图中分析出来;从(2)中可见,A 和 B 之间无建议关系,B 和 D 相互提供建议,其他情况都可从图中分析出来。



(1) 朋友关系图



(2) 建议关系图

	A	B	C	D	E
A	—	1	0	0	0
B	1	—	1	0	0
C	0	1	—	1	1
D	0	0	1	—	1
E	0	0	1	1	—

	A	B	C	D	E
A	—	0	1	1	1
B	0	—	0	1	1
C	1	0	—	0	1
D	1	1	0	—	0
E	1	1	1	0	—

图 12.5　朋友关系矩阵和建议关系矩阵

上述矩阵数据表达了 5 个人之间的朋友关系和建议关系。因此,各个观察值之间不相互独立,用许多标准的统计程序就不能对其进行参数估计和统计检

验,否则会计算出错误的标准差。对于这个问题,学者们利用一种随机化检验(randomization test)方法来检验,QAP 即属于这种方法之一。QAP 是一种以重新抽样为基础的方法,已经在社会网络研究中得到了广泛的应用,其研究对象都是"关系"数据。QAP 可以计算出正确的标准差(参见:http://fmwww.bc.edu/RePEc/nasug2001/simpson.pdf)。

QAP 要求进行的随机化操作步骤如下。为了比较两个矩阵之间的相关性,首先把每个矩阵中的所有取值看成是一个长向量,每个向量包含 $n(n-1)$ 个数字(对角线上的数字忽略不计)。然后像比较任何两个变量之间的相关性那样计算这两个向量之间的相关系数。把上述二个矩阵变成两个长向量,可计算出二者之间的相关系数(Krackhardt, 1987:176)为−0.8165,这表明二者之间是负相关的。也就是说,越是"朋友"关系,越不提供"建议"。问题是,这个观察到的相关系数值在统计意义上是否显著?实际上,我们不能利用标准的统计推断方法来研究这个问题,因为其前提性假设被违背。此时可以采用一种重排法(permutation approach)。我们要追问的实质性问题是,相对于在各个关系的取值是随机的情况下计算得到的相关系数来说,观察到的相关系数是大还是小?

我们可以随机地对矩阵的标签(A, B, C, D, E)进行置换,然后计算相关系数。就上述"朋友关系"和"建议关系"例子来说,我们可以对其中的一个矩阵(如建议关系矩阵)中的"标签"进行置换,对于包含 5 个行动者的矩阵来说,其最多可能有 5! =120 种置换方式。按照图 12.6 的置换方式计算出来的与"朋友矩阵"之间的相关系数为正数 0.8165;按照图 12.7 的置换方式计算出来与"朋友矩阵"之间的相关系数则为 0,这说明向谁寻求建议与"是否为朋友"无关。

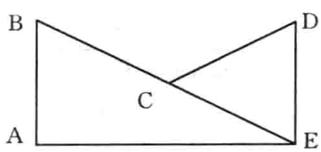


图 12.6 建议关系的第一种转置图

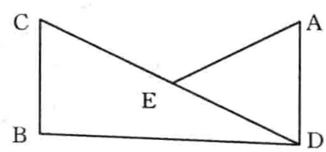


图 12.7 建议关系的第二种转置图

	A	B	C	D	E
A	—	1	0	0	1
B	1	—	1	0	0
C	0	1	—	1	1
D	0	0	1	—	1
E	1	0	1	1	—

建议关系的第一种转置图矩阵

	A	B	C	D	E
A	—	0	0	1	1
B	0	—	1	1	0
C	0	1	—	0	1
D	1	1	0	—	1
E	1	0	1	1	—

建议关系的第二种转置矩阵

表 12.17 全体 120 种置换的分布

相关系数值	置换的次数	百分比
0.8165	4	3.3
0.4082	32	26.7
0.0000	48	40.0
-0.4082	32	26.7
-0.8165	4	3.3
加权的均值:0.0	总数:120	100

由表 12.17 可见,在"建议关系矩阵"的全部 120 个置换矩阵中,有 4 个(占总数的 3.3%)与"朋友关系"是强相关的(相关系数为 0.816 5);有 32 个(占总数的 26.7%)与"朋友关系"是中等程度相关的(相关系数为 0.4082);有 48 个(占总数的 40%)与"朋友关系"是无关的,相关系数为 0.0;其余的 30%是负相关。从这种分析中可以得出如下结论:如果图 12.5 代表实际数据,那么这种实际结果表明这种置换只有 3.3%的概率是随机造成的。也就是说,如果原假设为"建议数据是随机指派的",那么在原假设成立的情况下,我们实际观察到的数据之间的关系出现了小概率事件(3.3%),因此,我们可以拒绝原假设,即认为这两种"关系"之间是存在负关系的。

该例子比较简单。如果网络规模较大,那么需要置换的次数将急剧增加。一些文献探讨了如何解决该问题(Mantel, 1967; Hubert & Schultz, 1976)(转引自 Krackhardt, 1987:176)。一个矩阵和另外一个矩阵的全部可能的转置矩阵之间的相关系数的均值和标准差恰恰是这两个矩阵元素值的函数。因此,观察到的相关系数可表示为一个标准化的 Z-值。假设在进行全部置换的情形下,置换的矩阵和原矩阵之间的相关系数服从标准正态分布的话,那么观察到的相关系数的显著性水平就仅仅取决于在标准正态曲线下超出 Z-值之外的区域面积(Krackhardt, 1987:178)。

上述案例是对数据标签进行重排。在具体计算中,QAP 对一个矩阵的行和列同时进行置换,具体做法有三步。

首先,计算已知的两个矩阵之间的相关系数。

其次,仅对其中的一个矩阵的行和相应的列同时进行随机的置换(而不是仅仅置换行或者列,否则破坏原始数据),然后计算置换后的矩阵与另一个矩阵之

间的相关系数[①],保存计算的结果;重复这种计算过程几百次甚至几千次,将得到一个相关系数的分布,从中可以看到这种随机置换后计算出来的几百或几千个相关系数大于或等于第一步中计算出来的观察到的相关系数比例。

最后,比较第一步中计算出来的实际观察到的相关系数与根据随机重排计算出来的相关系数的分布,看观察到的相关系数是落入拒绝域还是接受域,进而做出判断。也就是说,如果上述比例低于 0.05(假设研究者确定的显著性水平为 0.05),这在统计意义上表明所研究的两个矩阵之间存在强关系,或者说二者之间的相关系数不太可能是随机带来的。

可见,QAP 是一种非参数检验(non-parametric test)。QAP 只对某个矩阵的行和列同时进行置换,然后计算置换后的矩阵与另一个矩阵(即模式矩阵)之间的相关系数,这就保证了自变量矩阵和因变量矩阵在行和列上都相互依赖。最后,计算出显著性以及相关系数大于或者小于实际系数的概率。表 12.18 所示,是置换检验与常规检验之间的区别。

表 12.18　置换检验与常规统计检验之间的区别

置换检验	常规检验
检验关系变量之间的关系;不关注总体的分布;非参数检验	检验属性变量之间的关系,随机样本,总体为正态分布,参数检验;检验的结果可以推广到总体

QAP 研究的范围很广:如拥有共同董事的公司在股票市场上的行为是否类似?年龄相仿的人是否容易成为朋友?

在解释 QAP 检验结果的时候,心中要记住 QAP 原假设的实质。可以将 QAP 理解为检验在控制已知矩阵结构的基础上,通过变动另外一个矩阵中特定点的标签引起的两个矩阵结构上的差异。实际上,在所研究的矩阵之间往往存在着实质性的同构。QAP 趋于给出非显著性的结果。一般情况下,利用 QAP 检验两个变量图统计量的时候,不能获得一个单尾显著性水平;反之,QAP 下的显著性并不一定意味着观察到的结构关系与人们从诸如图的规模、密度等典型

① 但要注意的是,这里所说的相关是指两个矩阵(作为两个变量)之间的相关,给出的 Pearson 系数是两个矩阵之间的相关性系数。还要指出的是,矩阵中的值是 1 或者 0,而在统计分析的时候通常不用 Pearson 系数来分析分类变量之间的相关性。但是,考虑到虚拟变量的平均值和方差都是有意义的,因此可以把两个矩阵看成是两个“虚拟变量”,视之为“连续变量”,在这个意义上也可以做两个矩阵之间的相关分析。因此,上述 Pearson 系数就是这样计算出来的。

结构中期望的结构关系是相关联的。在研究诸如“朋友关系”和“建议关系”是否存在关系这样例子的时候，QAP 就是一种非常好的可以排除掉虚假结构关系的方法。

总之，QAP 是一种用来检验关系矩阵之间关系的方法。它可以计算出显著性水平，因而可用于对一些关系性命题进行检验。

二、QAP 相关分析

QAP 相关分析既可以研究两种“关系”矩阵之间是否相关，还可以研究一个属性(如性别)与一种关系(如朋友关系)是否有关。下面分别加以展示。

帕吉特等社会学者(Padgett & Ansell, 1993)从历史文献中收集到文艺复兴时期佛罗伦萨 16 个家族之间的社会关系数据，具体包括商业关系数据(PADGB)和婚姻关系数据(PADGM)。这两种二值对称数据保存在 UCINET 自带的名为 PADGETT 的数据文件中。

为了展示 QAP 的操作步骤，下面利用 UCINET 中的 QAP 命令计算“婚姻关系”和“商业关系”之间的相关性，从而对该假设进行检验。如果在统计意义上显著，则研究假设“‘婚姻关系’和‘商业关系’之间有关”成立。

在 UCINET 中，沿着 Tools→Testing Hypotheses→Dyadic(QAP)→QAP correlation 这条路径，点击后，出现一个对话框。然后，在“Matrix to Correlate”下面的空白处键入 PADGM(或者从磁盘中选出该文件)，回车之后，再键入 PADGB(或者从磁盘中选出该文件)①，点击 OK 后得到的运算结果给出了 PADGM(婚姻关系)和 PADGB(商业关系)之间的相关关系检验。在计算结果中，如果相关系数在统计意义上显著，则上述假设成立。

这里的“Obs Value”指的是在这两个矩阵之间实际观察到的相关系数，该值是 0.372；“Significa”指显著性水平，它等于 0，可见结果是很显著的；“Average”指的是根据 5 000 次随机置换计算出来的相关系数的均值，它等于 0.001；“Std Dev”是标准差 0.091；Minimum 代表随机计算的相关系数中出现的最小值；Maximum 代表随机计算的相关系数中出现的最大值 0.372；Prop>=0 表

---

① PADGM(婚姻关系)和 PADGB(商业关系)这两个矩阵并不单独存在于 UCINET 的数据库中，而是共同存在于 PADGETT 这个文件之中。因此，在计算之前，需要用 Data→unpack 程序先对 PADGETT 进行“开包”处理，进而把这两个矩阵打开，然后才能提交运算。

表 12.19 PADGM(婚姻关系)和 PADGB(商业关系)之间的相关关系检验

```
QAP CORRELATION
--------------------------------------------------------------------------------
Data Matrices:                 H:\whole UCINET\数据\PADGB
                               H:\whole UCINET\数据\PADGM
# of Permutations:             5000
Random seed:                   24233

QAP results for H:\whole UCINET\数据\PADGM * H:\whole UCINET\数据\PADGB (5000 permutations)

                      Obs Value Significa Average  Std Dev Minimum Maximum Prop>=0 Prop<=0
                      --------- --------- -------  ------- ------- ------- ------- -------
Pearson Correlation       0.372     0.000   0.001    0.091  -0.169   0.372   0.000   1.000

QAP Statistics

QAP Correlations
          PADGB PADGM
          ----- -----
    PADGB 1.000 0.372
    PADGM 0.372 1.000

QAP P-Values
          PADGB PADGM
          ----- -----
    PADGB 0.000 0.000
    PADGM 0.000 0.000
```

明这些随机计算出来的相关系数大于或等于实际相关系数的概率接近 0；Prop<=0 表明这些随机计算出来的相关系数小于或等于实际相关系数的概率，接近 1。

由表 12.19 可知，上述两个矩阵之间是正相关的，相关系数为 0.372，并且关系在统计的意义上是显著的。因此，“婚姻关系”与“商业关系”在统计意义上是相关联的。也就是说，在文艺复兴时期，这 16 个家族之间的“婚姻关系”与“商业关系”有关。

三、QAP 关系列联表分析

利用 Tools → Testing Hypotheses → Dyadic (QAP) → QAP Relational Crosstabs 这条路径，选择 UCINET 自带数据 KNOKBUR 中的“KNOKI”(诺克教授信息交换数据)和“KNOKM”(诺克教授的资金交换数据)作为输入矩阵，即可计算出这两个关系矩阵之间的关系列联表，如表 12.20 所示。

可见，对于任何两个点来说，二者之间的关系在“KNOKM”和“KNOKI”上对应的元素之值都是“0”的情况有 30 个；在“KNOKM”上取值为“0”，但是在“KNOKI”上取值为“1”的情况有 38 个；在“KNOKM”上取值为“1”，但是在“KNOKI”上取值为“0”的情况有 11 个；二者之间的关系在“KNOKM”和“KNOKI”上对应的元素值都是“1”的情况有 11 个。给出的这种交互表可以说明这两个矩阵之间的关系。

接下来给出的是这种列联表分析的统计量，是根据 2 000 次随机置换得到

表 12.20　矩阵之间关系列联表

```
RELATIONAL CROSSTABS
--------------------------------------------------------------------------------
Data Matrices:                H:\whole UCINET\数据\KNOKI
                             H:\whole UCINET\数据\KNOKM
# of Permutations:            2000
Random seed:                  24322
--------------------------------------------------------------------------------

Cross-Tab of KNOKM * KNOKI
                 0.000     1.000
             --------- ---------
    0.000           30        38
    1.000           11        11

Statistics for KNOKM * KNOKI (2000 permutations)
              Obs Value Significa Average  Std Dev Minimum Maximum Prop>=0 Prop<=0
              --------- --------- -------  ------- ------- ------- ------- -------
  Chi-Square      0.232     0.858   1.577    2.125   0.000  15.438   0.858   0.271
 Correlation     -0.051     0.410  -0.000    0.132  -0.414   0.365   0.719   0.410
----------------------------------------------------------------------------------
```

的。可见,无论是卡方统计量,还是相关系数,都不显著。这或许是因为数据规模较小,或许实际情况就是如此。

四、QAP 回归分析

QAP 回归分析的目的是研究多个矩阵和一个矩阵之间的回归关系,并且对判定系数 R^2 的显著性进行评价。在具体计算的时候要经过两步。首先,针对自变量矩阵和因变量矩阵对应的长向量元素进行常规的多元回归分析;其次,对因变量矩阵的各行和各列进行(同时)随机置换,然后重新计算回归,保存所有的系数值以及判定系数 R^2 值。重复这个步骤几百次,以便估计统计量的标准误(standard errors)。对于每个系数来说,该程序将计算出在全部随机置换的次数中,产生的系数大于或等于第一步计算时得到系数的随机置换所占的比例。在 QAP 多元回归时,要求所有变量(即所有矩阵)必须是 1-模矩阵,即必须是 N×N 的方阵。如果给出的是"个人—事件"关系网络等隶属关系数据,可以利用 UCINET 中的算法(Data→Affiliations(2-Mode to 1-Mode))转换为 N×N 矩阵。关于 QAP 多元回归分析,这里仅举 UCINET 自带的数据——24 个国家之间的多种关系数据加以说明(有关 QAP 回归的具体细节,请参见 Krackhardt,1988)。

在 UCINET 中,有一个自带数据名为 trade,它记载了 24 个国家之间的 5 种关系:外交关系、原材料贸易关系、食品贸易关系、制造品交易关系和矿物质贸易关系。首先利用 UCINET 中的 Data→Unpack 程序将该文件"开包"成为 5 个独立的文件,文件名分别是 DIPLOMATIC_EXCHANGE(外交关系)、CRUDE_

MATERIALS(原材料)、FOODS(食品)、MANUFACTURED_GOODS(制造品)和 MINERALS(矿物质)。下面展示的是利用 QAP 方法进行回归分析的结果。

点击 Tools→Testing Hypothesis→Dyadic(QAP)→QAP Regression→Double Dekker Semi-Partialling(MRQAP),在随后出现的对话框的“Dependent variable”一项中键入(或者从相应的地址找到)作为因变量的外交关系矩阵(DIPLOMATIC_EXCHANGE),在“Independent variables”一项中分别输入(或者从相关文件夹中选出)作为自变量的四个矩阵名 CRUDE_MATERIALS、FOODS、MANUFACTURED_GOODS 和 MINERALS,如图 12.8 所示。

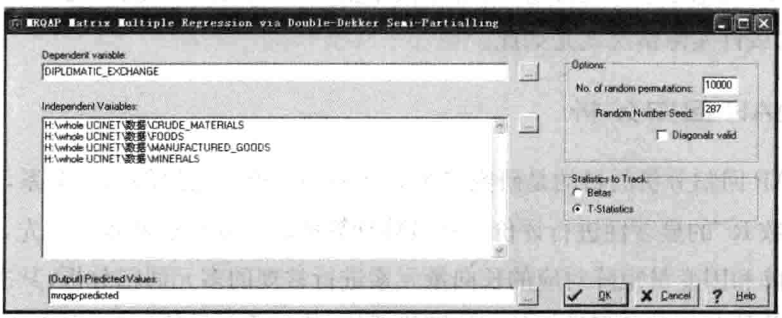


图 12.8 矩阵回归对话框

点击 OK 后得到的结果如表 12.21。

表 12.21 矩阵回归结果

```
MULTIPLE REGRESSION QAP VIA DOUBLE DEKKER SEMI-PARTIALLING
----------------------------------------------------------------------
# of permutations:              10000
Diagonal valid?                 NO
Random seed:                    348
Dependent variable:             H:\whole UCINET\数据\DIPLOMATIC_EXCHANGE
Expected values:                H:\whole UCINET\数据\mrqap-predicted
Independent variables:          H:\whole UCINET\数据\CRUDE_MATERIALS
                                H:\whole UCINET\数据\FOODS
                                H:\whole UCINET\数据\MANUFACTURED_GOODS
                                H:\whole UCINET\数据\MINERALS

Number of permutations performed: 10000

MODEL FIT

R-square     Adj R-Sqr     Probability       # of Obs
--------     ---------     -----------     -----------
   0.317         0.314           0.000             552
```

REGRESSION COEFFICIENTS

Independent	Un-stdized Coefficient	Stdized Coefficient	Significance	Proportion As Large	Proportion As Small
Intercept	0.339307	0.000000			
CRUDE_MATERIALS	0.109233	0.115284	0.011	0.011	0.989
FOODS	0.049976	0.052744	0.143	0.143	0.857
MANUFACTURED_GOODS	0.367435	0.387285	0.000	0.000	1.000
MINERALS	0.140151	0.127965	0.006	0.006	0.994

Running time: 00:00:08
Output generated: 09 十月 08 21:58:37
Copyright (c) 1999-2008 Analytic Technologies

分析结果主要包括两个部分。第一部分简单介绍模型的拟合情况，计算出来的确定系数为 0.317，调整的确定系数(Adj R-Sqr)为 0.314，应该说这个值是比较大的。这说明，当知道"原材料"(CRUDE_MATERIALS)、"食品"(FOODS)、"制造品"(MANUFACTURED_GOODS)和"矿物质"(MINERALS)等四个关系变量与"外交关系"变量之间存在"线性关系"的时候，可以用上述四个矩阵数据解释掉"外交关系"变异的 31.4%。判定系数右侧的一项是概率。要注意，这个概率是满足如下条件的随机置换的次数与总随机置换次数之比(也是一种概率)，即该次随机置换产生的判定系数不小于实际观察到的判定系数，它是单尾检验(One-Tailed Probability)所得的概率，本例中这个值接近于 0。

"# of Obs"这一项指的是观察项的数目，它等于 552。这是因为我们有 24 个国家构成的 24 行 24 列矩阵，其中当然有 24×(24－1)＝552 个观察项(一个国家与"自身"不存在贸易关系)。

第二部分给出了截距、每个自变量的非标准化回归系数、标准化回归系数和统计显著性检验的结果等。这里给出了两个概率(Proportion As Large 和 Proportion As Small)。要注意，与以上类似，这两个概率是满足如下条件的随机置换的次数与总随机置换次数之比，即在"Proportion As Large"情况下，随机置换产生的判定系数绝对值不小于观察到的判定系数；在"Proportion As Small"情况下，随机置换产生的判定系数的绝对值不大于观察到的判定系数。显然，此处进行的是双尾检验(Two-Tailed Probability)。

在本例中，三个自变量(CRUDE_MATERIALS、MANUFACTURED_GOODS 和 MINERALS)的回归系数在统计的意义上是"显著"的(其显著性水平分别是 0.011，0.000 和 0.006)，其中 MANUFACTURED_GOODS 最显著。而自变量"FOODS"不显著($p=0.143$)，这表明"食品"对"外交关系"没有贡献。

第四节 密度检验

还有一类假设检验问题涉及密度检验。该类问题又包含两个具体问题:(1)一个网络的密度是否等于某个值?(2)两个网络的密度是否相等?下面分别讨论。

一、一个网络的密度是否等于某个值

有时候我们关注已知的一个网络的密度是否等于研究者自己根据某种理论得到的值。例如,虚无假设可能是:密度为0,或者密度不等于0.5。对于多值矩阵来说,虚无假设还可能是 H_0:密度不是2.4。在此类情况下,要检验上述假设是否成立,可利用UCINET中的Network→Compare densities→Against theoretical parameter程序。

例如,对于组织之间的关系来说,我们假设所有的组织都有一种将信息直接传递给其他组织的趋势,从而维持自己的合法性。如果该命题成立,那么在UCINET软件中自带的诺克教授信息网络(命名为KNOKI)的密度值在理论上应该是1.0。当然,实际数据的密度是0.544,而不是1(沿着Network→Cohesion→Density→(new)Density Overall路径即可计算出其密度)。现在的问题是,实际的密度0.544与理论预测的密度1之差有可能是随机误差造成的,如果是随机造成的,就可以认为其密度值是1。由于我们不知道差值是不是随机造成的,因此对这个问题需要进行假设检验。点击Network→Compare densities→Against theoretical parameter,选择KNOKI数据,如图12.9所示。

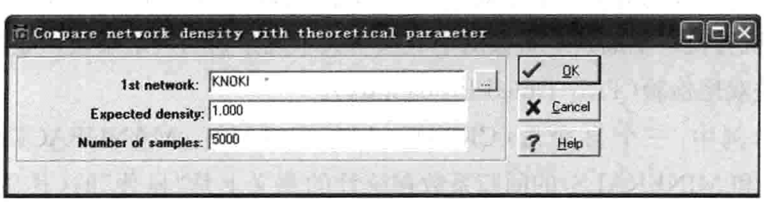


图 12.9 密度比较对话框

在"Expected density"中键入我们希望检验的密度值1。参数"Number of samples"通过从KNOKI数据中随机抽取5 000次样本,并计算密度均值的样本

方差，从而构造密度值的抽样分布，进而估计其标准误(standard error)。在统计学中，这个抽样过程叫做“自举抽样”(bootstrapping)。通过这种方法可以得到一个统计量的抽样分布，该分布的标准差就是标准误。点击 OK 后，得到的假设检验结果如表 12.22。

表 12.22　密度比较结果

```
COMPARE DENSITY W/ HYPOTHESIZED VALUE
--------------------------------------------------------------

Parameter value is: 1.0000
Density of KNOKI is: 0.5444
Difference is: -0.4556
Variance of ties for KNOKI: 0.2508
Classical estimate of SE: 0.0528

Number of bootstrap samples: 5000
Estimated standard error for density of KNOKI: 0.1201
z-score: -3.7925
Average bootstrap density: 0.4893
Proportion of absolute differences as large as observed: 1.0000
Proportion of differences as large as observed: 1.0000
Proportion of differences as small as observed: 0.0002
--------------------------------------------------------------
```

由此可见，需要检验的值是 1.000，实际观察值是 0.5444，二者之差是 −0.4556。如果原假设(密度等于 1)在总体中成立，那么该差值在多大程度上是由随机误差造成的？如果利用常规的均值标准误公式 $s/\sqrt{N}=\sqrt{0.2508}/\sqrt{90}$，可以计算出抽样的标准误估计值为 0.0528。用这个值进行检验，检验统计量就是 $-0.4556/0.0528=-8.629$，根据自由度为 $N-1=89$ 的 t 检验进行判断，显然这是一个具有统计显著意义的结果。

然而，这种统计检验是不可信的，因为这是“关系”数据。因此，还是要利用自举法，随机抽取出各个点的子集，构造 5 000 个网络并计算其密度，得到的抽样分布的均值是 0.4893，标准误是 0.1201。利用这种标准误重新计算，得到的 t 统计量为 −3.7943，该值当然还是显著的($p=0.0002$)。由于这里的原假设是“密度等于 1”，备择假设是“密度不等于 1”，既然计算结果显著，则拒绝原假设，接受备择假设，所以结论是“密度不等于 1”。

为什么利用常规统计方法得到的标准误的估计值 0.0528 远远小于利用自举法得到的值 0.1201？这仍然是因为常规的标准误公式有一个前提条件，即所有的观测值(即全部关系)相互独立。但是，由于关系本身实际上产生于同样的 10 个行动者，该前提条件是不满足的，因此，自举法更适用。通常情况下，标准的统计推断公式(例如标准误公式)常常给出偏小的估计结果，不符合实际情况，

据此进行统计推断可能出现错误。

二、两个网络的密度是否相等

计算表明,诺克教授的信息交换网络(KNOKI)密度高于资金交换网络(KNOKM)的密度。也就是说,同一群行动者之间的两种关系的密度却不同。问题是,这种差异是随机造成的,还是现实果真如此?这就需要利用另外一种假设检验程序计算。

利用 Network→Compare densities→Paired(same node)程序会完成这项任务,它将比较同一群行动者的两类关系之间的密度是否相同,计算估计的标准误,利用自举法检验差值是否显著。如果两类关系都是二值的,该检验相当于对一类关系中的概率与另外一类关系中的概率之差的检验。当两类关系都是多值关系的时候,该检验相当于对两类关系中的关系强度均值之差的检验。

下面仍然以诺克教授的信息交换网与资金网为例,进行这种检验。在 UCINET 中,点击 Network→Compare densities→Paired(same nodes)程序,在"1st network"中选择"KNOKI. ##h",在"2nd network"中选择"KNOKM. ##h",如图 12.10 所示(本次计算选择自举抽样的次数为 10 000 次)。

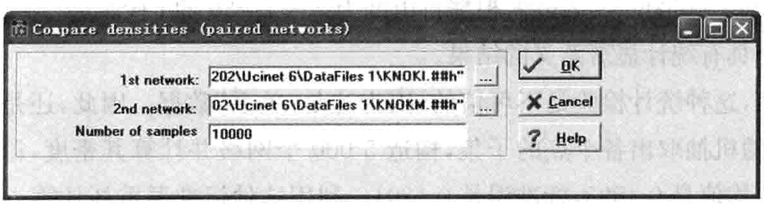


图 12.10　两个矩阵之间密度比较对话框

点击 OK 后,得到如表 12.23 所示的结果。

在计算结果中既给出了常规统计检验的结果,也给出了利用自举法检验的结果。均值(或者比例、密度)之差是 0.3000。利用常规法得到差值的抽样分布的标准误是 0.0697,而利用自举法计算得到差值的标准误是 0.1237。可见,常规法得到的结果明显小于自举法结果,这意味着常规法在拒绝原假设(即认为两个矩阵的密度相等)方面过于乐观了。

表 12.23　Knoke 信息交换数据与资金往来数据密度之差检验结果

```
BOOTSTRAP PAIRED SAMPLE T-TEST
--------------------------------------------------------------------------------

Density of "KNOKI.##h" is: 0.5444
Density of "KNOKM.##h" is: 0.2444
Difference in density is: 0.3000

Number of bootstrap samples: 10000
Variance of ties for "KNOKI.##h": 0.2508
Variance of ties for "KNOKM.##h": 0.1868
Classical standard error of difference: 0.0697
Classical t-test (indep samples): 4.3024
Estimated bootstrap standard error for density of "KNOKI.##h": 0.0965
Estimated bootstrap standard error for density of "KNOKM.##h": 0.0775
Bootstrap standard error of the difference (indep samples): 0.1237
95% confidence interval for the difference (indep samples): [0.0575, 0.5425]
bootstrap t-statistic (indep samples): 2.4247
Bootstrap SE for the difference (paired samples): 0.1259
95% bootstrap CI for the difference (paired samples): [0.0531, 0.5469]
t-statistic: 2.3820
Average bootstrap difference: 0.2587
Proportion of absolute differences as large as observed: 0.0178
Proportion of differences as large as observed: 0.0052
Proportion of differences as small as observed: 0.9949
```

利用自举法得到的结果是，双尾检验概率为 0.0178。而如果我们事先知道备择假设中差值的方向，就可以利用单尾检验结果 $p=0.0052$，这个值显然很小。因此，我们可以很确信地得出结论，这些组织之间的信息交换关系的密度在统计意义上显著地大于资金关系的密度。也就是说，观察到的两个矩阵密度之差来自随机误差的概率极低，或者说，实际情况的确如此。

本 章 小 结

常规统计方法要求变量之间要独立，而关系数据恰恰违背这个假设。因此，在检验"关系"之间关系的时候往往不能用常规的统计方法。本章介绍了三类假设检验：点层次的假设检验，包括回归分析、t 检验和方差分析；点—关系混合层次的假设检验，包括"类别变量与关系变量之间关系的检验"与"连续变量与关系变量之间关系的检验"；关系—关系层次的假设检验，包括 QAP 相关分析，QAP 关系列联表分析和 QAP 回归分析。本章介绍了 QAP 的原理，并用实际例子展示 QAP 的应用。本章重在介绍以置换为基础的各种检验方法，它不同于常规统计检验，因而更加适用于检验关系数据之间的关系。最后简要介绍了密度检验方法。

参 考 文 献

[美]布劳，1991/1977，《不平等和异质性》，王春光等译，北京：中国社会科学出版社。

[美]林顿·C.弗里曼，2008/2004，《社会网络分析发展史》，张文宏、刘军、王卫东译，北京：中国人民大学出版社。

郭志刚，1999，《社会统计分析方法——SPSS 软件应用》，北京：中国人民大学出版社。

[英]吉登斯，1998/1984，《社会的构成》，李康、李猛译，北京：生活·读书·新知三联书店。

林南，2005，《社会资本》，张磊译，上海：上海人民出版社。

刘军，2004，《社会网络分析导论》，北京：社会科学文献出版社。

刘军，2006a，《法村社会支持网络——一个整体研究的视角》，北京：社会科学文献出版社。

刘军，2006b，《和谐社会的形式基础——三方关系研究》，载《中国社会学会 2005 年学术年会优秀论文集》，北京：社会科学文献出版社，第 136—148 页。

刘军，2007，《一般化互惠：测量、动力及方法论意含》，《社会学研究》第 1 期。

刘军，2008，《社会网络模型及其哲学意义研究》，哈尔滨：黑龙江大学博士后工作报告。

刘军、杨辉，2012，《从实体论到关系论——关系社会学的认识论》，《北方论丛》第 6 期。

刘军、David Willer、Pamela Emanuelson，2013a，《强制结构理论与实验检验》，《社会》第 4 期。

刘军、郭丽娜，2013b，《关系对联盟的影响》，《社会学研究》第 6 期。

刘小枫、苏国勋主编，2005，《社会理论的诸理论》，上海：上海三联书店，第 458—488 页。

[英]罗德里克·马丁著，1992，《权力社会学》，丰子义、张宁译，北京：生活·

读书·新知三联书店。

罗家德,2002,《NQ 风暴——关系管理的智慧》,北京:社会科学文献出版社。

罗家德,2010,《社会网分析讲义》(第二版),北京:社会科学文献出版社。

时蓉华,2004/1998,《社会心理学》,杭州:浙江人民出版社。

[英]斯科特,2007/2000,《社会网络分析法》,刘军译,重庆:重庆大学出版社。

[美]斯坦利·沃瑟曼、凯瑟琳·福斯特,2012/1994,《社会网络分析:方法与应用》,陈禹、孙彩虹译,北京:中国人民大学出版社。

[荷]沃特·德·诺伊、安德烈·姆尔瓦、弗拉迪米尔·巴塔盖尔吉,2012,《蜘蛛:社会网络分析技术》,林枫译,北京:世界图书出版公司。

吴金闪、狄增如,2004,《从统计物理学看复杂网络研究》,《物理学进展》第1期。

谢宇,2006,《社会学方法与定量研究》,北京:社会科学文献出版社。

张文宏,2006,《中国城市的阶层结构与社会网络》,上海:上海人民出版社。

周雪光,2003:《组织社会学十讲》,北京:社会科学文献出版社。

周怡,2000,《社会结构:由"形构"到"解构"》,《社会学研究》第3期。

Abend, Gabriel, 2008, "The meaning of 'theory'." *Sociological Theory* 26, 173—199.

Alba R.D., and Kadushin, C., 1976, "The intersection of social circle: A new measurement of social proximity in network." *Sociological Methods and Research* 5, 77—102.

Alba, R.D., 1973, "A graph-theoretic definition of a sociometric clique." *Journal of Mathematical Sociology* 3, 113—126.

Alba, R.D., and Moore, G., 1978, "Elite social circle." *Sociological Methods and Research* 7, 167—188.

Alexander, S., "Using the bipartite line graph to visualize 2-mode social networks." In http://www.casos.cs.cmu.edu/events/conferences/2005/2005_proceedings/Alexander.pdf.

Bell, D.C., J.S. Atkinson, and J.W. Carlson, 1999, "Centrality measures

for disease transmission networks." *Social Networks* 21, 1—21.

Berkowitz, S. D., 1982, *An Introduction to Structural Analysis: The Network Approach to Social Research*. Torento: Butterworths.

Bernard, HR., Killworth, PD., & Sailor, L., 1982, "Informant accuracy in social network data V: An experimental attempt to predict actual communication from recall data." *Social Science Research* 11:30—66.

Bonacich, P., Lloyd., P., 2001, "Eigenvector-like measures of centrality for asymmetric relations." Downloaded from Bonacich's webpage. (http://www.sscnet.ucla.edu/soc/faculty/bonacich/home.htm).

Bonacich, P. B., 1987, "Power and centrality: A family of measures." *American Journal of Sociology* 92, 1170—1182.

Bonacich., P., Cody, A., Michael., H.February 18, 2002, "Hyper-edges and multi-dimensional." Download from Bonacich's webpage.

Bonanich, P., Holdren, A.C., and Johnston, M., 2004, "Hyper-edges and multidimensional centrality." *Social Networks* 26, 189—203.

Borgatti, S.P. & Foster, P.C., 2003. "The network paradigm in organizational research: A review and typology." *Journal of Management* 29, 991—1013.

Borgatti S. P., & Everett, M. G., 1997, "Network analysis of 2-mode data." *Social Networks* 19, 243—269.

Borgatti, S.P., & Everett, M.G., 1989, "The class of all regular equivalences: Algebraic structure and computation." *Social Networks* 11, 65—88.

Borgatti, S.P., & Everett, M.G., 1992a, "Graph colorings and power in experimental exchange networks." *Social Networks* 14, 287—308.

Borgatti, S.P., & Everett, M.G., 1992b, "Notions of position in social network analysis." *Sociological Methodology* 22, 1—35.

Borgatti, S.P., & Everett, M.G., 1992c, "Regular blockmodels of multiway, multimode matrices." *Social Networks* 14, 91—120.

Borgatti, S.P., & Everett, M.G., 1993, "Two algorithms for computing regular equivalence." *Social Networks* 15, 361—376.

Borgatti, S.P., & Everett, M.G., 1999, "Models of core/periphery struc-

tures." *Social Network* 21, 375—395.

Borgatti, S.P., & Everett, M.G., 2006, "A graph-theoretic perspective on centrality." *Social Networks* 28, 466—484.

Borgatti, S.P., & Everett, M.G., and Shirey, P.R., 1990, "LS sets, lamda sets, and other cohesive subsets." *Social Networks* 12, 337—358. Down from Borgatti's personal webpage.

Borgatti, S.P., 1997, "Consensus analysis." http://www.analytictech.com/borgatti/consensu.htm.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2014. *UCINET for Windows: Software for Social Network Analysis*. Harvard: Analytic Technologies.

Borgatti, S.P., Everett, M.G. and Johnson, J.C., 2013, *Analyzing Social Networks*. Sage Publications UK.

Boyd, J.P., and Jonas, K.J., 2001, "Are social equivalence ever regular? Permutation and Exact Tests." *Social Networks* 23, 87—123.

Brandes, U., 2008, "On variants of shortest-path betweenness centrality and their generic computation." *Social Networks* 30, 136—145.

Braun, N., 1997, "A rational choice model of network status." *Social Networks* 19, 129—142.

Breiger, R.L., 1974, "The duality of persons and groups." *Social Forces* 53, 181—190. Also in Wellman and Berkowitz, eds., 1988, *Social Structure: A Network Approach*. Cambridge: Cambridge University Press, 83—98.

Breiger, R.L., 1976, "Career attributes and network structure: A block-model study of a biomedical research specially." *American Sociological Review* 41, 117—135.

Breiger, R.L., 2003, "Emergrnt themes in social network analysis: Results, challenges and opportunities." In Breiger, R.L.Kathleen Carley, and Philippa Pattison. Eds. 2003. *Dynamic Social Network Modeling and Analysis*. The National Academic Press.

Burt. R.S., 1976, "Positions in networks." *Social Forces* 55, 93—122.

Burt, R.S., 1982, *Toward a Structural Theory of Action*. London: Aca-

demic Press.

Burt, R. S., 1992, *Structural Holes: the Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Burt, R.S., 2004, "Structural holes and good ideas." *American Journal of Sociology* 110, 349—399.

Caulkins, D., and Susan B. Hyatt, 1999, "Using consensus analysis to measure cultural diversity in organizations and social movements." *Field Methods* 11, 5—26.

Caulkins, D., 2004, "Identifying culture as a threshold of shared knowledge: A consensus analysis method." *International Journal of Cross Cultural Management* 4, 317—333.

Coleman, J., Katz, E., and Menzel, H., 1957, "The diffusion of an innovation among physicians." *Sociometry* 20, 253—270.

Collins, R., 1988, *Theoretical Sociology*. New York: Harcourt Brace Jjovanovich.

Costenbader, E., Valente, T.W., 2003, "The stability of centrality measures when networks are sampled." *Social Networks* 25, 283—307.

Cynthia M. Webster, Anita L. Iannucci, and A. Kimball Romney., 2002, "Consensus analysis for the measurement and validation of personality traits." *Field Methods* 14, 46—64.

Davis, A., Gardner, B.B. and Gardner, M.R., 1941, *Deep South*. Chicago: University of Chicago Press.

Degenne, A., & Forse, M., 1999, *Introducing Social Networks*. London: Sage Publications.

Doreian P., 1974, "On the connectivity of social networks." *Journal of Mathematical Sociology* 3, 245—258.

Doreian, P., 1969, "A note on the detection of cliques in valued graphs." *Sociometry* 32, 237—242.

Doreian, P., Batagelj, V., and Ferligoj, A., 2000, "Symmetric-acyclic decompositions of networks." *Journal of Classification* 17, 3—28.

Doreian, P., Batagelj, V., and Ferligoj, A., 2005, *Generalized Block-*

modelling. New York: Cambridge University Press.

Doreian, P., Batagelj, V., and Ferligoj, A., 2004, "Generalized blockmodeling of two-mode network data." *Social Networks* 26, 29—53.

Doreian, P., and A. Mevar., 2009, "Partioning signed social networks." *Social Networks* 31, 1—11.

Dunbar, R.I.M., 1992, "Neocortex size as a constraint on group size in primates." *Journal of Human Evolution* 22, 469—984.

Emirbayer, Mustafa and Jeff Goodwin, 1994, "Network analysis, culture, and problem of agency." *American Journal of Sociology* 99, 1411—1454.

Erdös, P., Rényi, A., 1959, "On random graphs. I." *Publicationes Mathematicae* 6, 290—297.

Erickson, B., 1988, "The relational base of attitudes." In Wellman, B., and Berkowitz, S.D.(eds.). *Social Structures:A Network Approach*, 99—121. Cambridge, England: Cambridge University Press.

Everett, M., 2002, *Social Network Analysis*. Essex: Textbook at Essex Summer School in SSDA.

Everett, M.G. and Borgatti, S.P., 1999, "The centrality of groups and classes." *Journal of Mathematical Sociology* 23, 181—202.

Fararo, T. J., and J. Skvoretz, 1987, "Unification research programs: Integrating two structural theories." *American Journal of Sociology* 92, 1183—1209.

Fararo, T.J., 1981, "Biased networks and social structure theorems. Part 1." *Social Networks* 3, 137—159.

Fararo, T.J., 1983, "Biased networks and the strength of weak ties." *Social Networks* 5, 1—11.

Fararo, T.J., and Skvoretz, J., 1984, "Biased networks and social structure theorems. Part 2." *Social Networks* 6, 223—258.

Faust, K., 1997, "Centrality in affiliation networks." *Social Networks* 19, 157—191.

Frank & Yasumoto, 1998, "Action and structure." *American Journal of*

Sociology 104, 642—686.

Frank, O., 1981, "A survey of statistical methods for graph analysis." In Leinhardt, S. ed., *Sociological Methodology*, 110—155. San Francisco: Jessey-Bass.

Freeman, L., 1992, "The sociological concept of group: An empirical test of two models." *American Journal of Sociology* 98, 152—166.

Freeman, L., 2004, *The Development of Social Network Analysis*. Vancouver: Empirical Press.

Freeman, L.C., 1977, "A set of measures of centrality based on betweenness." *Sociometry* 40, 35—41.

Freeman, L.C., 1979, "Centrality in social networks: Conceptual clarification", *Social Networks* 1, 215—239.

Freeman, L.C., 1992, "The sociological concept of 'group': An empirical test of two models." *American Journal of Sociology* 98, 55—79.

Freeman, L. C., 1972, "Segregation in social networks." *Sociological Methods and Research* 1, 411—430.

Freeman, L.C., S.P.Borgatti and D.R.White, 1991, "Centrality in valued graphs: A measure of betweenness based on network flow." *Social Networks* 13, 141—154.

Friedkin, N.E., 1984, "Structural equivalence and cohesion explanations of social homogeneity." *Sociological Methods and Research* 12.

Friedkin, N.E. and E.C. Johnsen, 1997, "Social positions in influence networks." *Social Networks* 19, 209—222.

Friedkin, N.E., 1991, "Theoretical Foundations for centrality measures." *American Journal of Sociology* 96, 1478—1504.

Friedkin, N. E., 1998, *A Structural Theory of Social Influence*. New York: Cambridge University Press.

Garfield, E., 1979, *Citing Indexing: Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley.

Gilbert, E.N., 1959, Random Graphs. *Annals of Mathematical Statistics* 30:1141—1144.

Gil-Mendieta, J., and Schmidt, S., 1996, "The political network in Mexico." *Social Networks* 18, 355—381.

Gould, J. and Fernandez, J., 1989, "Structures of mediation: A formal approach to brokerage in transaction networks." *Sociological Methodology*. 89—126.

Gould, R., 1991, "Multiple networks and mobilization in the paris commune, 1871." *American Sociological Review* 56, 716—729.

Granovetter, M., 1973, "The strength of weak ties." *American Journal of Sociology* 78, 1287—1303.

Grant, K.L., and Miller, M.L., 2004, "A cultural consensus analysis of marine ecological knowledge in the Solomon Islands." SPC Traditional Marine Resouse Management and Knowledge Information Bulletin # 17—December 2004.

Griffith, D.A., 1987, "Spatial autocorrelation: A primer." Association of American Geographers, Resource Publications in Geography.

Heider, F., 1946, "Attitudes and cognitive organization." *Journal of Psychology* 21:107—112.

Heider, F., 1958, *The Psychology of Interpersonal Relation*. John Wiley & Sons.

Hummon, Norman P., Doreian, Patrick, 2003, "Some dynamic of social balance processes: Bringing heider back into balance theory." *Social Network 25*:17—49. Elsevier Science.

Hanneman, R.A., 2008, *Social Network Analysis*. http://faculty.ucr.edu/~hanneman/nettext/.

Hanneman, Robert A. and Mark Riddle, 2005, *Introduction to Social Network Methods*. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/).

Hubbell, C.H., 1965, "An input-output approach to clique identification." *Sociometry* 28, 377—399.

Hummon, N. P., and Doreian, P., 2003, "Some dynamics of social balance processes: Bringing heider back into balance theory." *Social Networks*

25, 17—49.

Jaskyte, Kristina and William W. Dressler, 2004, "Studying culture as an integral aggregate variable: organizational culture and innovation in a group of nonprofit organizations." *Field Methods* 16, 265—284.

Kadushin, C., 1966, "The friends and supporters of psychotherapy: On social circles in urban life." *American Sociological Review* 31, 786—802.

Kadushin, C., 1968, "Power, influence and social circles: A new methodology for studying opinion makers." *American Sociological Review* 33, 685—699.

Kadushin, C., 2012, *Understanding Social Networks*. New York: Oxford University Press.

Katz, L., 1953, "A new status index derived from sociometric data analysis." *Psychometrika* 18, 34—43.

Khanafiah, Deni and Hokky Situngkir, 2011, "Social balance theory: Revisiting heider's balance theory for many agents." in http://cogprints.org/3641/.

Knoke, D., and Kuklinski, J. H., 1982, *Network Analysis*. Beverly Hills, Calif.: Sage.

Krackhardt, D., 1987, "QAP partialling as a test of spuriousness." *Social Networks* 9, 171—186.

Krackhardt, D., 1988, "Predicting with networks: Non-parametric multiple regression analysis of dyadic data." *Social Networks* 10, 171—186.

Krackhardt, David and Robert N. Stern, 1988, "Informal networks and organizational crises: an experimental simulation." *Social Psychology Quarterly* 51, 123—140.

Krackhardt, David, 1994, "Graph theoretical dimensions of informal organizations." In Kathleen Carley & Michael Prietula (eds.). *Computational Organizational Theory*. Hillsdale, NJ: Lawrence Erlbaum Associaties, Inc, 89—111.

Levine, Joel H., 1972, "The sphere of influence." *American Sociological Review* 37, 14—27.

Lorrain, Françoise and H.C.White, 1971, "Structural equivalence of indi-

viduals in social networks". *Journal of Mathematical Sociology* 1, 49—80.

Lusher, Dean, Johan Koskinen, and Garry Robbins, 2012, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. London: Cambridge University Press.

Marc L. Miller, John Kaneko, Paul Bartram, Joe Marks, and Devon D. Brewer, 2004, "Cultural consensus analysis and environmental anthropology: Yellowfin tuna fishery management in Hawaii." *Cross-Cultural Research* 38, 289—314.

Marsden, P., 1984, "Measuring tie strength." *Social Forces* 63, 482—501.

Marsden, P., 1990, "Network data and measurement." *Annual Review of Sociology* 16, 435—463.

Marsden, P., 2002, "Egocentric and sociocentric measures of network centrality." *Social Networks* 24, 407—422.

Milgram, S., and Travers, J., 1969, "An experimental study of the small world problem." *Sociometry* 32, 425—443.

Mitchell, J. C., 1969, "The concept and use of social network." In Mitchell(ed.), *Social Networks in Urban Settings*. Manchester: Manchester University Press.

Mizruchi, M. S., 1994, "Social network analysis: Recent achievements and current controversies." *Acta Sociologica* 37, 329—343.

Mizruchi, M. S. and B. B. Potts, 1998, "Centrality and power revisited: Actor success in group decision making." *Social Networks* 20, 353—387.

Moody, J., 2004, "The structure of a social science collaboration network." *American Sociological Review* 69, 213—238.

Moody, J., 2008, Moody 的个人主页: http://www.soc.duke.edu/~jmoody77/s884/index.htm.

National Research Council, 2003, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, Ronald Breiger, Kathleen Carley, and Philippa Pattison. Committee on Human Factors. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nowicki, K. & Snijders, T.A.B., 2001, "Estimation and prediction for stochastic block models."*Journal of the American Statistical Association* 96, 1077—1087.

Oliver, A. L., Yuval Kalish and Gad Yair, 2007, " Reflection on 'brokerage and closure'." *Social Networks* 29, 330—339.

Padgett, J.F. and C.K.Ansell, 1993, "Robust action and the rise of the medici, 1400—1434." *American Journal of Sociology* 98, 1259—1319.

Peng Wang, Ken Sharpeb, Garry L.Robins, Philippa E.Pattisona, 2009, "Exponential random graph (p*) models for affiliation networks." *Social Networks* 31, 12—25.

Poulin, R., M.C.Boily and B.R.Masse, 2000, "Dynamical systems to define centrality in social networks." *Social Networks* 22, 187—220.

Robin, G., P.Elliot, & P.Pattison, 2001, "Network models for social selection processes." *Social Networks* 23, 1—30.

Robin, G., Pattison, P., & P.Elliot, 2001, "Network models for social influence processes." *Psychometrika* 66, 161—190.

Robin, G., Pattison, P., & Woolcock, J., 2004, "Missing data in networks: exponential random graph models (p*) for networks with non-respondents." *Social Networks* 26, 187—220.

Robin, G., Pattison, P.E. & Wasserman, S., 1999, "Logic models and logistic regression for social networks, Ⅲ, valued relations." *Psychometrika* 64, 371—394.

Rosenthal, N., Fingrutd, M., Ethier, M., Karant, R., McDonald, D., 1985, "Social movements and network analysis: A case study of nineteenth-century women' reform in New York State." *American Journal of Sociology* 90, 1022—1054.

Scott, J., 2013, *Social Network Analysis: A Handbook*. (3rd edition) London: Sage Publications.

Carrington, Peter J., John Scott, and Stanley Wasserman, (eds.) 2005, *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.

Seidman, S.B., and B.L., Foster, 1978, "A graph-theoretic generalization of the clique concept." *Journal of Mathematical Sociology* 6, 139—154.

Simmel, G., 1902, "The number of members as determining the sociological form of the group." *American Journal of Sociology* 8, 1—46; 8. 158—196.

Simpson, W., *QAP.*, http://fmwww. bc. edu/RePEc/nasug2001/simpson.pdf.

Skvoretz, Fararo and Agneessens, 2004, "Advances in biased net theory: Definitions, derivations, and estimations." *Social Networks* 26, 113—139.

Skvoretz, J., and Faust, K., 1999, "Logit models for affiliation networks." *Sociological methodology* 29, 253—280.

Snijders, T. A. B. and Borgatti, S. P., 1999, "Non-parametric standard errors and tests for network statistics." Connections, 22(2):61—70.(从 Borgatti 的个人主页:http://www.analytictech.com/borgatti/中下载.)

Snijders, Tom A. B. Philippa E. Pattison, Garry L. Robins, and Mark S.Handcock, 2006, "New specifications for exponential random graph models." *Sociological Methodology* 36, 99—153.

Snijders, Tom A.B.Gerhard G. van de Buntb, Christian E.G.Steglich, 2010. "Introduction to stochastic actor-based models for network dynamics." *Social Networks* 32, 44—60.

Snyder, D., and Kick, E.L., 1979, "Structural position in the world system and economic grouth, 1955—1970: A multiple-network analysis of transnational interactions." *American Journal of Sociology* 84, 1096—1126.

Taylor, M., 1969, "Influence structures." *Sociometry* 32, 490—502.

Travers, J, Milgram, S., 1969, "Anexperimental study of the small world problem." *Sociometry* 32, 425—443.

Wasserman, S. & Pattison, P. E., 1996, "Logic models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*." *Psychometrika* 60, 406—426.

Wasserman, S. and K.Faust, 1994, *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserman, S., and Anderson, C., 1987, "Stochastic a posteriori block-

models: Construction and assessment." *Social Networks* 9, 1—37.

Watts, Duncan. J., 2003, *Six Degrees: The Science of a Connected Age*. New York: W.W.Norton & Company, Inc.

Watts, D., 1999, "Networks, dynamics and the small-world phenomenon." *American Journal of Sociology* 105, 493—527.

Watts, D., 1999, *Small Worlds*. Princeton University Press, Princeton, New Jersey.

Wellman, B., and Berkowitz, S.D., 1988, (eds.). *Social Structures: A Network Approach*. Cambridge, England: Cambridge University Press.

White, H.C., Boorman, S.A., and Breiger, R.L., 1976, "Social structure from multiple networks. I. blockmodels of roles and positions." *American Journal of Sociology* 81, 730—779.

Willer, D., 1999, *Network Exchange Theory*. Praeger Publishers, Westport CT.

William W.Dressler, 1996, "Culture and blood pressure: Using consensus analysis to create a measurement." *Field Methods* 8, 6—8.

Yasuda Y. and H.Tokuraku, 1999, "On recursive definition of action and structure." *Sociological Theory and Methods* 13, 183—192.

后　　记

社会网络分析是对社会关系进行量化研究的一门社会学分支学科，该学科在国外已经有80多年的发展历史。相比之下，我国的社会网络研究仍然比较薄弱，尽管中国目前仍然是一个重视“关系”的国家。近30多年来，社会网络分析在整体网模型方面又取得重大突破。整体网数据如何收集、分析和整理？整体网有哪些指标可以测量？如何应用UCINET软件分析网络数据？这些内容是进行整体网研究的基础，希望本书对国内的社会网络研究尽绵薄之力。“社会网络分析”领域的重要文献都来自国外，本书的撰写也离不开这些文献。笔者所引用的著作、文章或讲义（如汉尼曼、穆迪、艾弗雷特、薄加提等教授的讲义）都标明了出处，在此向相关作者表示感谢！笔者做的工作就是梳理文献，结合自己的理解展示UCINET软件的应用。介绍他人的“著述”，再“编”上自己的一点读书体会，这种“科研成果”或许就是“编著”。

在“编著”过程中，英国爱丁堡大学的马丁·艾弗雷特教授，美国波士顿大学的斯蒂芬·薄加提教授，加州大学河边分校的罗伯特·汉尼曼教授，杜克大学的詹姆斯·穆迪教授解答了我的部分疑问。清华大学的罗家德教授拨冗赐序，哈尔滨工程大学校长刘志刚教授、副书记高晚欣教授全力支持社会学事业，人文学院院长金宏章教授、副院长郑莉教授发自内心地支持笔者的研究工作，此缘难见，本书的修订工作也离不开责任编辑王亚丽的不断“督促”。笔者一并表示衷心的感谢！

整体网研究的内容十分广泛。本书只简单介绍了其中的一些内容与方法，还有一些内容没有涉及。如果本书能够促进读者进一步思考并研究整体网，笔者的目标就达到了。“人无完人，金无足赤”，笔者想以此为借口来掩饰书中的不足。本书不当甚至错漏之处，读者可发信（liujunry@163.com）告之，笔者将十分感谢。

刘　军

2014年6月　哈尔滨盟科观邸寓所

世纪出版

社会网络分析被认为是与理性选择学派、新制度理论学派并举的重要学派，在管理学、社会学等领域有很大影响，而国内相关论著比较少。本书梳理了大量国内、国外研究成果，全面介绍了社会网分析的重要内容——整体网分析，以及如何使用UCINET软件实施测量。本书最大的特点是全面、详尽、实用，是读者使用UCINET软件进行整体网分析的得力助手。

格致出版社向**使用本书授课的教师**提供教学资源
授课教师可以通过如下方式联系到我们：
邮箱：hibooks@hibooks.cn

微信　　微博

责任编辑：王亚丽
封面设计：人马艺术设计·储平

上架建议：管理学·社会学
ISBN 978-7-5432-2398-1
9 787543 223981 >
定价：48.00元
易文网：www.ewen.cc
格致网：www.hibooks.cn

封面
书名
版权
前言
目录
第一章　社会网络分析简介
第一节　社会网络分析的含义
一、社会网络的含义
二、1-模网和2-模网
第二节　用图和矩阵表示社会网络
一、社会网络的图形表达
二、社会网络的矩阵表达
第三节　社会网络研究的范围和理论贡献
一、研究范围
二、理论贡献
第二章　整体网研究概要
第一节　整体网研究的内容
一、整体网的分类
二、整体网的构成和规模
三、整体网的密度
四、整体网成员之间的距离
五、整体网的结构研究
第二节　整体网资料的收集
一、整体网资料的收集方法
二、问卷设计应注意的问题
三、整体网资料的整理和分析
四、整体网资料的相关问题
第三节　整体网研究的优势及局限
一、整体网研究的优势
二、整体网研究的局限
第三章　整体网分析工具——UCINET简介
第一节　UCINET的运行环境
一、UCINET的数据分析格式
二、UCINET的主菜单功能简介
三、UCINET中矩阵运算的句法
第二节　UCINET数据的输入
一、初始数据文件

二、Excel文件数据
三、数据语言类型的文件
四、UCINET数据表编辑器
第三节　UCINET数据的预处理
一、抽取子图和子矩阵
二、数据的合并
三、置换与排序
四、转置和改型
五、重新编码
六、线性转换
七、对称化处理
八、捷径距离和可达性
九、矩阵的组合
十、标准化处理
十一、改变模态
第四章　均衡性分析
第一节　符号图中的均衡研究
一、社会心理学中的均衡研究
二、有关均衡的模拟研究
三、整体均衡研究
第二节　有向关系中的均衡
一、有向P-O-X三角关系
二、正—负关系的均衡性分析
三、有向三方组的谱系
第五章　中心性和影响力
第一节　度数中心性
一、点的度数中心度
二、图的度数中心势
第二节　中间中心性
一、点的中间中心度
二、图的中间中心势
三、线的中间中心度
四、针对中间中心度的等级嵌套分析
第三节　接近中心性
一、点的接近中心度
二、图的接近中心势
三、多种中心度指数的同时计算

第四节　影响力分析
一、影响力的含义及测量
二、影响力指数计算示例
第五节　群体中心度
一、三类群体中心度指数的含义
二、群体中心度指数的计算举例
第六章　凝聚子群分析
第一节　凝聚子群概说
一、社会结构的质性分析和量化研究
二、凝聚子群的类型
第二节　基于互惠性的凝聚子群
一、派系的定义
二、对派系概念的评价
第三节　基于可达性的凝聚子群
一、n-派系
二、n-宗派
第四节　基于度数的凝聚子群
一、k-丛
二、k-核
第五节　基于“子群内外关系”的凝聚子群
一、成分
二、块
三、LS集合
四、Lambda集合
五、社会圈
第六节　凝聚子群的分派指数
一、单类网络中的分派指数
二、多类网络中的分派指数
第七节　凝聚子群分析示例
一、进行凝聚子群分析的几个指导性步骤
二、凝聚子群分析示例
第七章　社会网络的关联性与凝聚度
第一节　关联度和凝聚力
一、根据社会关系界定的凝聚力
二、根据独立途径数界定的凝聚力
第二节　关联度的测量
一、关联度

二、图的等级度
三、图的效率
四、最近上限
第三节　关联性与小世界
一、小世界的含义
二、小世界的测量指标
三、小世界的测量示例
第八章　块模型分析
第一节　块模型的含义
一、块模型
二、CONCOR方法
第二节　块模型的应用举例
第九章　个体网研究：结构洞与中间人
第一节　局部桥和结构洞
一、局部桥的含义及测量
二、结构洞
第二节　中间人分析
一、中间人的分类
二、中间人次数的计算
第三节　其他个体网指标的计算
一、个体网的指标
二、个体网的同类性
三、个体网的构成
四、诚实中间人指数
第十章　一致性分析
第一节　一致性分析的含义
第二节　一致性分析举例
第十一章　2-模网研究
第一节　2-模数据的矩阵分析和图形分析
一、2-模网络向1-模矩阵的转化
二、二部“2-模图”分析
第二节　2-模数据的定量分析
一、2-模数据的中心性分析
二、2-模数据的奇异值分解
三、2-模数据的因子分析
四、2-模网络的对应分析
第三节　2-模数据的核心—边缘分析和分派分析

一、2-模数据的核心—边缘分析
二、2-模数据的分派分析
第十二章　关系数据的假设检验
第一节　点的属性层次的假设检验
一、回归分析
二、t检验
三、方差分析
第二节　点—关系混合层次的假设检验
一、类别变量与关系变量之间关系的检验
二、连续变量与关系变量之间关系的检验
第三节　关系—关系层次的假设检验
一、QAP的含义
二、QAP相关分析
三、QAP关系列联表分析
四、QAP回归分析
第四节　密度检验
一、一个网络的密度是否等于某个值
二、两个网络的密度是否相等
参考文献
后记
封底